

**UUM NETWORK TRAFFIC ANALYSIS AND USER' WEBSITE  
PREFERENCES**

**MUSTAFA MOHAMMED IBRAHIM AL-KAWAZ**

**UNIVERSITI UTARA MALAYSIA  
2012**

**UUM NETWORK TRAFFIC ANALYSIS AND USER' WEBSITE  
PREFERENCES**

**BY**

**Mustafa Mohammed Haki Ibrahim**

**(808988)**



KOLEJ SASTERA DAN SAINS  
(College of Arts and Sciences)  
Universiti Utara Malaysia

**PERAKUAN KERJA KERTAS PROJEK**  
(Certificate of Project Paper)

Saya, yang bertandatangan, memprakukan bahawa  
(I, the undersigned, certifies that)

**MUSTAFA MOHAMMED IBRAHIM AL-KAWAZ**  
**(808988)**

calon untuk Ijazah  
(candidate for the degree of) **MSc. (Information Technology)**

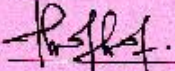
telah mengemukakan kertas projek yang bertajuk  
(has presented his/her project of the following title)

**UUM NETWORK TRAFFIC ANALYSIS AND USERS' WEBSITE PREFERENCES**


seperti yang tercatat di muka surat tajuk dan kulit kertas projek  
(as it appears on the title page and front cover of project)

bahawa kertas projek tersebut telah diterima dari segi bentuk serta kandungan  
dan meliputi bidang ilmu dengan memuaskan.  
(that this project is in acceptable form and content, and that a satisfactory  
knowledge of the field is covered by the project).

Nama Penyelia  
(Name of Supervisor) : **DR. MOHD HASBULLAH OMAR**

Tandatangan  
(Signature) :  Tarikh (Date) : 27/6/2012

Nama Penyelia  
(Name of Supervisor) : **MR. ADIB M. MONZER HABBAL**

Tandatangan  
(Signature) :  Tarikh (Date) : 27/6/2012

DEAN OF AWANG HAD SALLEH GRADUATE SCHOOL  
UNIVERSITI UTARA MALAYSIA

**PERMISSION TO USE**

In presenting this study in partial fulfilment of the requirements for a postgraduate degree from the Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this study in any manner in whole or in part, for scholarly purposes may be granted by my supervisor(s) or in their absence by the Dean of Awang Had Salleh Graduate School. It is understood that any copying or publication or use of this study or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my study.

Requests for permission to copy or to make other use of materials in this study, in whole or in part, should be addressed to

Dean of Awang Had Salleh Graduate School  
College of Arts and Sciences  
Universiti Utara Malaysia  
06010 UUM Sintok  
Kedah Darul Aman  
Malaysia

## **Abstract**

The current world is experiencing a revolution in Internet services and networking; a revolution that provided and continues to provide varying features and invaluable tools to computer networks. On the other hand, several problems are being faced by users and global organizations in networking including lack of bandwidth and packet loss during transmission which impacts Internet efficiency and the performance of network. These issues can be rectified through the measurement and analysis of the network's performance. Moreover, for network performance enhancement, it is imperative to study users' behaviour. Therefore, the main objectives of the present study are to identify UUM network performance through Internet traffic and to highlight users' behaviour. A total of three methodological steps are carried out to meet the objectives of the study; the first one is the data collection phase whereby the source for data collection is taken from the presently used main distributed switch in an hour for each working day in a duration of one week; the second one is the data analysis phase where Wireshark is used to provide the statistics of traffic and finally; the third phase is the data presentation where Microsoft Excel is utilized to present data. Study findings presents valuable information of network bandwidth, data loss rates and Ethernet traffic distribution, in addition to the fact that is social websites are the most websites used in UUM. These findings leads to facilitate the enhancement of UUM network performance and Internet bandwidth strategies.

## Acknowledgement

*First and foremost, Alhamdulillah, All praise is to my Lord, the Compassionate, and the Merciful Subhanahhuwata 'alah; for giving me the will and strength in the completion of this study.*

*I would like to express my deepest gratitude and appreciation to my respective Supervisor: Dr. Mohd. Hasbullah bin Omar and to my second supervisor: Mr. Adib M. Monzer Habbal and Mr. Khuzairi bin Mohd Zaini for their expertise, kindness, and patience in guiding throughout the production of this Study.*

*My excessive gratefulness goes to Dr. Mohammed Haki, my spiritual mentor. The first, last and always, a lasting heartfelt gratitude to the source of my light and pleasure, to the one who enlightens my life, to my dear Mother. Equal gratitude goes out to my precious Sister.*

*Finally I am also thankful to the people I met in my life who touched my heart and gave me strength to move forward to something better, Mr Adli for his assistance on data collection at the computer centre. My dear brother Houzifa Mohammed Hentaya and my friend Abed-Alsalam Tayara.*

Mustafa Mohammed Haki AL-Kawaz

June 15, 2012

## Table of Contents

Abstract.....	ii
Acknowledgement .....	iii
Table of Contents.....	iv
List of Tables .....	vii
List of Figures .....	viii
List of Appendices .....	x
List of Abbreviations .....	xi
<b>CHAPTER ONE INTRODUCTION .....</b>	<b>1</b>
1.1 Introduction .....	1
1.2 Background .....	3
1.3 Problem Statement .....	5
1.4 Objective .....	5
1.5 Project Questions .....	6
1.6 Scope of the Study .....	6
1.7 Organization of the Project .....	6
1.8 Summary .....	7
<b>CHAPTER TWO LITERATURE REVIEW .....</b>	<b>8</b>
2.1 Introduction .....	8
2.2 Network Performance .....	8
2.2.1 Network Throughput.....	9
2.2.2 Network Measuring Tools .....	11
2.2.2.1 Tcpcdump Program .....	12
2.2.2.2 Wireshark Program.....	13
2.2.3 Ethernet Protocols .....	14
2.2.3.1 Internet Protocol Version Six (IPv6).....	15
2.2.3.2 Internet Protocol Version Four (IPv4).....	15
2.3 Users Preference and Websites Categories .....	20
2.3.1 Social Networks .....	22
2.3.2 Blogs .....	22
2.3.3 E-commerce and Services .....	23

2.4 Related Work .....	23
<b>CHAPTER THREE PROJECT METHODOLOGY .....</b>	<b>27</b>
3.1 Introduction .....	27
3.2 Data Collection.....	28
3.2.1 Devices and Topology .....	28
3.2.2 Packets Capture.....	31
3.3 Data Analyzing.....	32
3.3.1 Wireshark Program .....	33
3.3.2 Using Wireshark to Achieve the First Objective .....	33
3.3.3 Using Wireshark to Achieve the Second Objective.....	35
3.4 Data Representation .....	37
<b>CHAPTER FOUR NETWORK PERFORMANCE .....</b>	<b>39</b>
4.1 Introduction .....	39
4.2 Network Performance .....	39
4.2.1 Network Load and Throughput Measuring.....	40
4.2.2 Packets Loss .....	42
4.2.3 Packets length distribution .....	42
4.3 Protocols Distribution .....	44
4.3.1 Ethernet Traffic Distribution.....	45
4.3.2 IPv4 Distribution.....	46
4.4 Summary .....	57
<b>CHAPTER FIVE INTERNET USERS PREFERENCES .....</b>	<b>58</b>
5.1 Introduction .....	58
5.2 Users' Preferred Category of Websites.....	58
5.3 Websites Preferences .....	60
5.3.1 Social Websites Preferences .....	60
5.3.2 Blogs Preferences.....	61
5.3.3 E-commerce and Technical Support Websites Preferences.....	62
5.3.4 News Websites Preferences .....	63
5.3.5 File Shearing Websites Preferences.....	64
5.3.6 Multimedia Websites Preferences.....	65



5.3.7 E-mail Websites Preferences .....	65
5.3.8 Search Engines Websites Preferences.....	66
5.3.9 Educational Websites Preferences .....	67
5.3.10 Informational and Services Websites Preferences .....	67
5.4 Countries Traffic Distributions .....	68
5.5 Summary .....	69
<b>CHAPTER SIX FINDINGS AND FUTURE RECOMMENDATION .....</b>	<b>70</b>
6.1 Introduction .....	70
6.2 Discussion of Findings .....	70
6.3 Suggestion of Future Works .....	71
6.4 Problems and Limitation .....	72
6.5 Contribution .....	72
<b>REFERENCES.....</b>	<b>74</b>
<b>APPENDIX A WEBSITES STATISTICS .....</b>	<b>81</b>

## **List of Tables**

Table 3.1: Size of Captured Data	32
Table 3.2: Wireshark Filters	34
Table 3.3: HTTP Filters	35
Table 4.1: All Day's Statistic	39
Table 4.2: Packets Loss	42
Table 4.3: Packets Length	43
Table 4.4: Ethernet Average Packets Size	46
Table 4.5: IPv4 Average Packets Size	50

## **List of Figures**

Figure 2.1: Network Performance Measurement Tools	12
Figure 2.2: Basic Tcpdump Commands	13
Figure 2.3: Ethernet Networks Topology	14
Figure 3.1: Project Methodology	27
Figure 3.2: UUM Network Topology	28
Figure 3.3: Capturing Device	29
Figure 3.4: Port Mirroring	30
Figure 3.5: Port Mirroring Codes	31
Figure 3.6: Wireshark IP Filters	36
Figure 3.6: OSPF Flow on Monday	38
Figure 4.1: Throughput of All Days	40
Figure 4.2: Network Load within Hour of Each Day	41
Figure 4.3: Packets Length for All Days	44
Figure 4.4: Protocols Analysis Stages	44
Figure 4.5: Ethernet Traffic Distribution	45
Figure 4.6: OSPF, ARP and GREP Packets Numbers	47
Figure 4.7: TCP, UDP and ICMP Packets Number	49
Figure 4.8: IPv4 Distribution	51
Figure 4.9: TCP Distribution	52
Figure 4.10: Con... TCP Distribution	53
Figure 4.11: UDP Distribution	55
Figure 4.12: Con... UDP Distribution	56
Figure 5.1: Websites Category User's Preferences	59
Figure 5.2: Websites Category User's Preferences Histogram	59
Figure 5.3: Social Networking Sites Distribution	62
Figure 5.4: Blogging websites Distribution	63
Figure 5.5: E-commerce and Technical Support Websites Packets	65
Figure 5.6: News Websites Packets	66
Figure 5.7: File Shearing Websites Packets	67

Figure 5.8: Multimedia Websites Packets Rates	68
Figure 5.9: E-mail Websites Packets Rates	69
Figure 5.10: Search Engines Websites Packets Rates	70
Figure 5.11: Educational Engines Websites Packets Rates	71
Figure 5.12: Informational & Services packets Rates	72
Figure 5.13: Packets Distributions over Countries	74
Figure 5.14: Local & International Packets Rates	75

## **List of Appendices**

Table 1: Social Websites Packets	80
Table 2: Blogging Websites Packets	80
Table 3: E-commerce and Technical Support Websites Packets	81
Table 4: News Websites Packets	82
Table 5: File Shearing Websites Packets	82
Table 6: Multimedia Websites Packets	83
Table 7: E-mail Websites Packets	83
Table 8: Search Engines Websites Packets	83
Table 9: Educational Websites Packets	84
Table 10: Informational & Services Websites Packets	84
Table 11: Packets Destinations	85

## **List of Abbreviations**

ARP	Address Resolution Protocol
DEC/RPC	Distributed Computing Environment / Remote Procedure Calls
DNS	Domain Name Service
FTP	File Transfer Protocol
GIF	Graphics Interchange Format
GREP	Generic Routing Encapsulation Protocol
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
IP	Internet Protocol
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
JPEG	Joint Photographic Experts Group
MIME	Multipurpose Internet Mail Extensions

OSPF	Open Shortest Path First
P2P	Peer To Peer
RTMP	Real Time Message Protocol
RTSP	Real Time Streaming Protocol
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
SSH	Secure Shell Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UUM	University Utara Malaysia
VLAN	Virtual Local Area Networks
YMG	Yahoo Messenger Protocol

# **CHAPTER ONE**

## **INTRODUCTION**

### **1.1 Introduction**

The Phenomenal success of the Internet has led to the rapid adoption of the Internet protocol technology to build all types of communication networks, including private corporate networks (intranet), military communication networks, home networks, and the emerging Third-generation cellular networks. Billions of devices worldwide are IP-capable, which allows remote access and control through the Internet. Such rapid and unprecedented convergence of communication through IP presents a host of challenging problems in guaranteeing the required performance in such networks (Jain & Hassan, 2004).

For the monitoring and security of network, it is imperative to acknowledge and expound on how the applications function. Many researchers have concentrated on the characteristics of traffic and network behavior under particular applications including P2P applications (Cao, Liu, & Xue, 2010).

In other words, a network may be defined as a “set of devices (often referred to as nodes) connected by communication links that are built using different physical media” (Marsic, 2010). A node can be represented by a computer, telephone or any device that facilitates the sending and receiving of messages while the medium of communication is referred to as the physical path through which the message flows from sender to receiver. Examples of media are fiber-optic cable, copper wire or air carrying radio waves (Marsic, 2010).



Network traffic measurement is the basis of network protocol design, equipment development, traffic engineering and network performance improvement. With the constantly expand of network bandwidth in recent years, several of network applications are growing faster and faster, network traffic growth in non-linear such as HTTP, FTP, TELNET and other traditional application are tending to be decreased even they have been reduced to below 1/3. About 2/3 flows are occupied by IP-based network games, audio/video, especially on P2P-based real-time streaming media, Internet phone, document sharing and other new applications. However, the extensive using of P2P software consumed a lot of bandwidth, resulted in network congestion and performance degradation (Chuan & Hong, 2008).

Another phase of network measurement is traffic or packet analysis, often referred to as packet sniffing or protocol analysis, describes the process of capturing and interpreting live data as it flows across a network in order to better understand what is happening on that network. Packet analysis is typically performed by a packet sniffer; a tool used to capture raw network data going across the wire. Packet analysis can help us understand network characteristics, learn who is on a network, determine who or what is utilizing available bandwidth, identify peak network usage times, identify possible attacks or malicious activity, and find unsecured and bloated applications (Sanders, 2007).

There are various types of packet sniffing programs, including both free and commercial ones. Each program is designed with different goals in mind. A few of the more popular packet analysis programs are Tcpcdump (a command-line program), OmniPeek, and Wireshark, both are GUI-based sniffers (COMER, 2009).

Following the great improvement of Internet services in universities and higher institutes of research and studies, it becomes imperative to determine where network resources are being utilized and where Internet traffic flows for the purpose of creating a strategy that improves the networks' efficiency.

UUM provides Internet access for over 28,000 of its students and 6,000 staff members. Moreover, the campus is linked to the Internet through TM-ISP company, the Internet provider. Every college in UUM is incorporated with a multilayer switch connecting to main switches in the computer center. The computer center in UUM comprises varying high-efficiency network devices including servers, firewall, controllers, multilayer switches and routers.

## **1.2 Background**

In a highly competitive network's world, network administrators are searching for the best performance that their network can perform. Capability and integrity of the network, these two elements cannot be achieved unless the network administrator knows exactly how all the components of their network work efficiently together, such as switches, routers and firewalls. (Blum, 2003). In networks, there are three major components should be measured to detect the performance of applications or web traffic that running over it, first is to monitor and analyze network bandwidth also called throughput "The bandwidth of a network is given by the number of bits that can be transmitted over the network in a certain period of time sometime called data rate" (Peterson & Davie, 2012). Second, latency sometimes called delay, in simple word latency is how much time the message (packet) takes to travel from one

host to another (Peterson & Davie, 2012). Third, Internet traffic characterization is an important topic for performance evaluation of networks (Hassan, Garcia, & Bockstal, 2009).

Multipath flow control has been proposed as a key way to improve the Internet performance, reliability, and flexibility in supporting changing loads. However, at this point, there are very few tools to quantify the performance benefits; particularly, in the context of a stochastic network supporting best-effort flows, e.g., file transfers and web browsing sessions (Joseph & Veciana, 2011).

Understanding the structure and dynamics of the virtual networks formed by Internet users and applications has become a major focus of Internet-related research. While these networks are of great sociological interest understanding their properties is also important for research topics as varied as intrusion detection, application design, and network capacity planning. However, this broad applicability creates a tension in the Internet community: researchers in many areas want to mine network data, but the primary data sources used by most analysis systems—captured packets and network flow data—are vast and contain personal and sensitive information (Meiss, Menczer, & Vespignani, March 2011).

### **1.3 Problem Statement**

The never ending demand for Internet services and network resources in the UUM campus resulted to some issues in monitoring network performance and activity management especially following 2003. Where new ideas of information exchange and distribution arose including Really Simple Syndication (RSS) and wikis which normally negatively impact network performance (Governor, Hinchcliffe, & Nickull, 2009). On top of this, novel applications were introduced which works under Internet services like peer to peer applications (Cao et al., 2010). Along with the increases in Internet traffic, particularly the increase in real-time applications, careful network planning and optimization are called for (Wamser, Pries, & Staehle, 2010). These factors among others result in high-traffic loads on network devices which often lead to the Internet users' dissatisfaction regarding networking speed and Internet browser's slow working although UUM has already enhanced the Internet speed. Hence, this study attempts to gain an in-depth understanding of the reason behind the phenomenon.

### **1.4 Objective**

The current study seeks to achieve the following objectives:

- To investigate of network traffic in UUM network in light of packet loss, throughput and network load.
- To identify users' preferred websites in relation to Internet traffic types in different periods of time.

## **1.5 Project Questions**

The present study aim to answer the following project questions:

- What is the condition of UUM network performance through an hour in a particular week?
- How does changing of users' preferred website impact the network performance during a certain period? What are the most frequently visited websites in the campus?

## **1.6 Scope of the Study**

This study comprises of five stages of observation and collection of network data in a period of one week starting from the second month of the initial semester. The statistics is carried out on the core multilayer switch at the computer center in UUM Universiti main campus which is already connected to the distributed switches and other various switches. A variety of capturing switches, analysis and presentation tools operate under Linux open-source operating system through lapcap and Round-Robin Database (RRD) library. Data capturing was only for an hour of each day, from 10:00 to 11:00 AM.

## **1.7 Organization of the Project**

This report is presented in six chapters .An overview of the content of the following chapters is as follows:

- Chapter 2: reviews the literature of this study in term of network performance measurement and websites user's preferences. Furthermore, provides

definitions of certain types of protocols and its prevalence. Overall the chapter highlights on the importance of studying network traffic.

- Chapter 3: describes the methodology used to analyze and measure the data, in order to achieve the study objectives.
- Chapter 4: presents the statistics and the results of network measuring process.
- Chapter 5: presents the statistics and the results of specifying websites user's preferences.
- Chapter 6: concludes the findings of the study.

## **1.8 Summary**

The nature of this study is expressed with clear objectives specifying particular research questions. The UUM campus network constitutes the course of the study. The network's performance is measured and analyzed over specific time duration in an attempt to determine the traffic type and the most frequently visited websites and to eventually provide the reasons leading to slow Internet browsing.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

The present chapter concentrates on reviewing prior literature and discussing the core subject of the study which is the examination of the network performance and users' preferences through of traffic collection and analysis. The chapter makes use of devices definitions which are germane to achieving the objectives of the study. Section 2.2 of the chapter, contains reviews of network performance variables, software and protocols that were previous used and studied in literature and their relation to the present study. The chapter explains certain devices functions such as routers, switches, and capturing device. Section 2.3 explains the Internet users' preferences and website types while section 2.4 expounds on related works.

#### **2.2 Network Performance**

Networks continue to change to support new applications, improve reliability and performance and reduce the operational cost. The changes are made to the network in the form of upgrades such as software or hardware upgrades, new network or service features and network configuration changes (Argyraki, Maniatis, & Singla, 2010). It is crucial to monitor the network when upgrades are made because they can have a significant impact on network performance and if not monitored may lead to unexpected consequences in operational networks. This can be achieved manually for a small number of devices, but does no scale to substantial networks with

hundreds or thousands of routers an extremely large number of different upgrades made on a regular basis (Mahimkar, Song, Ge, & Shaikh, 2010,).

Networks are at the core of global communication. The impact network changes have on network performance has been of concern to researchers. In various configurations, network performance has been evaluated using many methods (Narayan, Lutui, & Vijayakumar, 2010). Previous study focused on analyzing and measuring the performance of IPv4 and IPv6 protocols. The next section describes the network measurement variables.

### **2.2.1 Network Throughput**

Effective throughput or bandwidth simply defined by the number of application byte transferred in a second, for large file transfers, the effective throughput of the application, also called the speed or bandwidth of the connection, is a key performance measure. An inefficient protocols algorithm or implementation can significantly reduce the effective throughput even if the underlying network provides a very high speed communication channel. Observed bandwidth can be different for forward and reverse paths (Jain & Hassan, 2004).

Observed throughput can vary over time. Throughput variation is a metric to measure the variability on the received bandwidth over given time scale. In general, the larger time scale, the lower the throughput variability. For a given context, it is important to define a time scale over which throughput variability should be measured (Jain & Hassan, 2004).



- **Packet Loss Rate:** It is the ratio of the number of correctly received packets at the destination to the total number of packets transmitted at the source (Chang, Huang, Lin, & Chuah, 2011). Major sources of packets loss are (1) buffer overflow at the intermediate router (2) packet corruption caused by transmission errors. A high packet loss can severely degrade the performance of data and multimedia applications (Jain & Hassan, 2004).

**Network devices:** This section illustrates all networking devices that have been mentioned through this study, and afterward captured tools will be discussed.

- **Routers:** Are the most well-known networking device, routers work on the network layer OSI model. Its main function is to connect numerous networks. It transmits data between different networks by observing the addresses of the network contained in the packets that's route. (Curtis & Taylor, 2005). To do so, routers use forwarding and routing functions, forwarding means once the packet arrives to the receiver port of the router, the router duty is to transfer the packet to the right output link. Routing algorithms would define the path along which packets flow from the sending host to the receiving host (Team, 2006). In general, a router is a stand-alone device, but it can carry out as software running on a host.
- **Switches:** in simple words, switch is a network device that transfers packets between hosts on the local network LAN, and sometimes it can work as a router (multilayer switch), common perception switches are working on data-link layer dealing with Mac address it cannot understand IP address (Peterson

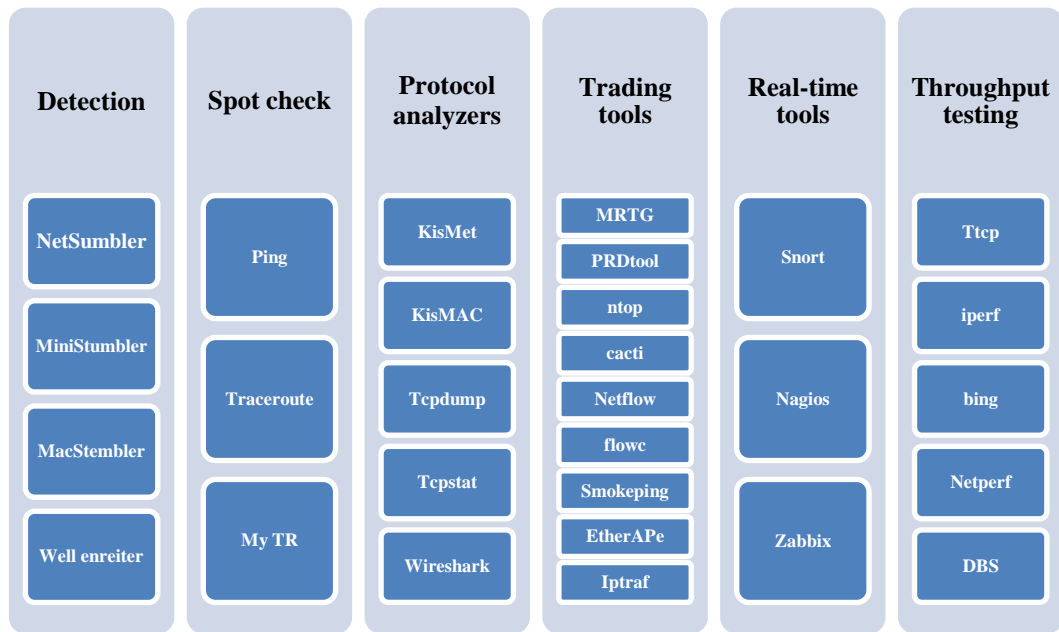
& Davie, 2012). Public implementations of network switches are Ethernet implementations.

- **Capturing device:** It could be a special device for packet capturing or using high specification computers such as servers. In case of using special device like Small-Net-builder to capture the data software will be implemented, on the contrary, in case of using a normal computer (Lucas, 2010).

### 2.2.2 Network Measuring Tools

Network performance measurement and prediction considers of the most prominent and indispensable components in distributed computing environments. The selection of the most advantageous network measurement tool or system for specific needs can be very time-consuming and may require detailed experimental analysis. The multi-dimensional aspects and properties of such systems or tools should be considered in parallel. (Yildirim, Suslu, & Kosar, 2009). To confirm that all network traffic will be captured network interface card (NIC) must be in promiscuous mode. Packet capturing libraries have been created in a UNIX environment to provide a common application programming interface (API) for packet capturing.

Libpcap is a free, open-source packet capture library originally developed at the Lawrence Berkeley National Laboratory in California (Orebaugh, Ramirez, & Burke, 2007). Libpcap library is used to capture the packets on the network directly from the network adapter. This library is an inbuilt feature of the operating system. In UNIX like systems pcap is implemented in the libpcap library (Qadeer & Khan, 2010). Following Figure demonstrates the most well-known tools:



*Figure 2.1: Network Performance Measurement Tools (Sloan, 2001)*

### 2.2.2.1 Tcpdump Program

The tcpdump program was developed at the Lawrence Berkeley Laboratory at the University of California, Berkeley, by Van Jacobson, Craig Leres, and Steven McCanne (Hartpence, 2011). It was originally developed to analyze TCP/IP performance problems. A number of features have been added over time, although some options may not be available with every implementation. The program has been ported to a wide variety of systems and comes preinstalled on many systems (Sloan, 2001).

Tcpdump is a command-line tool for monitoring network traffic it can capture and display all network protocol information down to the link layer. It can show all the packets header and data received, or just the packets that match particular criteria

(Jain & Hassan, 2004). Following Figure describes the most important Tcpdump commands:

Tcpdump commands	Description
-a	Attempt to convert network and broadcast addresses to names.
-c	Exit after receiving count packets.
-C	File size If the file is larger than the file size, close the current save file and open a new one.
-dd	Dump packet-matching code as a C program fragment.
-ddd	Dump packet-matching code as decimal numbers (preceded with a count).
-e	Print the link-level header on each dump line.
-E algo:secret	Use algo:secret for decrypting IPsec ESP packets.
-f	Print foreign Internet addresses numerically
-F	-F file Use file as input for the filter expression.
-i interface	Listen on interface.
-l	Make stdout line buffered.
-m module	Load SMI MIB module definitions from file module.

*Figure 2.2 : Basic Tcpdump Commands (Jain & Hassan, 2004)*

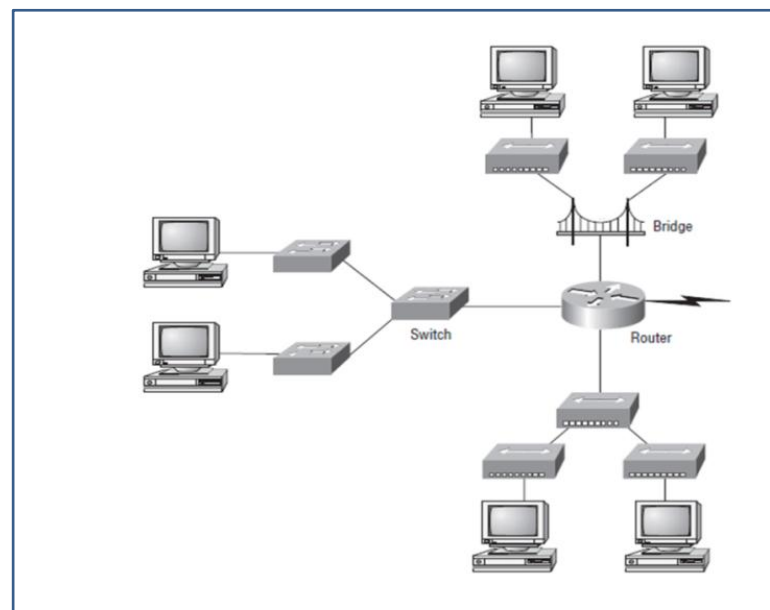
#### **2.2.2.2 Wireshark Program**

Formerly known as Ethereal, Wireshark is a free network protocol analyzer for Unix and Windows. Wireshark main work is examining data from a live network or from a capture file on disk, and interactively browses and sorts the captured data. Both summary and detailed information is available for each packet, including the full header and data portions. “ Wireshark has several powerful features, including a rich display filter language and the ability to view the reconstructed stream of a TCP, UDP, FTP and other session” (Orebaugh et al., 2007).

Network administrators use it to troubleshoot network problems, network security engineers use it to examine security problems, developers use it to debug protocol implementations, students use it to learn the TCP / IP protocol (Xu, Wang, & Yan, 2010).

### 2.2.3 Ethernet Protocols

An Ethernet network “is a segmented LAN technology that uses Ethernet NICs and twisted pair, coax, or fibre media. Ethernet is used to connect computers, printers, and servers within the same building or campus” (Curtis & Taylor, 2005).



*Figure 2.3: Ethernet Networks Topology (Lammle, 2007)*

### **2.2.3.1 Internet Protocol Version Six (IPv6)**

It was developed primarily to address the rapidly shrinking supply of IPv4 addresses, but also to implement some novel features based on the experience with IPv4. It is much simpler than the IPv4 header, but also double the size of a default IPv4 header, mainly because of longer IP addresses in IPv6, IPv6 address space is 128-bits (Marsic, 2010).

Moreover, IPv6 allows three types of address's Unicast, any-cast and Multicast. IPv6 has been on the way for more than 10 years now, yet for much of the world. It has been irrelevant until recently. Now, as the shortage of IPv4 addresses begins to become obvious to even the most hardened skeptic, awareness and interest are growing (McFarland, Sambhi, Sharma, & Hooda, 2011).

### **2.2.3.2 Internet Protocol Version Four (IPv4)**

Addresses was standardized to be 32-bits long, which gives a total of four billion possible network addresses (Marsic, 2010). IPv4 addresses are expected to completely run out by the end of 2012. Some websites are already facing address exhaustion today. With IPv6 in sight or already deployed, adding network address translation as a third addressing scheme can be avoided by introducing IPv6-only nodes (Knoth & Neuhäuser, 2010). In Chapter Four certain IPv4 protocols were mentioned, it will be explained in the following:

### **Open Shortest Path First (OSPF)**

OSPF is a link-state protocol that uses flooding of link-state information and Dijkstra least path algorithm. With OSPF routers constructs a complete topological map of the entire autonomous system. Router broadcasts routing information to all other routers in the autonomous system, not just to its neighbor routers, some advantages of this routing protocol are 1) Security: exchanges between OSPF routers can be authenticated, 2) Multi same-cost paths: OSPF allows multiple paths to be used when it has the same cost, 3) Integrated support for Unicast and Multicast routing, and 4) support for hierarchy within a single routing domain (Kurose & Ross, 2010).

### **Address Resolution Protocol (ARP)**

ARP Is a “telecommunications protocol interrogates the local network by sending out a broadcast asking the machine with the specified IP address to reply with its hardware address” (Lammle, 2007). Because there are both network layer address (IP address) and link layer-layer address (MAC) there is a need to translate between them, this is the job of the ARP protocol.

### **Internet Control Message Protocol (ICMP)**

ICMP is used by network devices to interconnect network-layer data between these devices. The most typical use of ICMP is for error reporting. This protocol often considered part of IP, but architecturally it lies just above IP, as ICMP messages are carried inside IP datagram. That is, ICMP messages are carried as IP payload, just as TCP or UDP segments are carried as IP payload (Kurose & Ross, 2010).

### **Generic Routing Encapsulation (GREP)**

This protocol is “Cisco-proprietary tunneling protocol. It forms virtual point-to-point links, allowing for a variety of protocols to be encapsulated in IP tunnels” (Lammle, 2007). GREP generating a virtual point-to-point connection to Cisco routers across an IP network at remote points. Furthermore, the most common networking protocols are TCP and UDP Protocols.

### **Transmission Control Protocol/Internet Protocol (TCP/IP)**

This protocol was considered as the most commonly using protocols in the Internet (Kouvatsos, 2011). “Transmission Control Protocol/Internet Protocol (TCP/IP) is a nonproprietary, routable network protocol suite that enables computers to communicate over all types of networks” (Curtis & Taylor, 2005).

By building on the functionality provided by the Internet Protocol (IP), the transport protocols delivers data to applications executing in the Internet this is done by making use of ports. The transport protocols provide additional functionality unlike User Datagram Protocol (UDP), such as congestion control, reliable data delivery, duplicate data suppression, and flow control as is done by the TCP (Parziale & Britt, 2006). Widely used TCP-applications include, electronic mail, Worlds Wide Web, File transfer and remote login. During this study, there were a variety of protocols that was used TCP to transfer the data. The following discusses major types of these protocols:



## **Hypertext Transfer Protocol (HTTP)**

Is one of the data-transfer protocols, HTTP is the web application-layer protocol, is at the heart of the web, implemented in two programs: A client program and the server program, these two programs execute on different end systems, communicating by exchanging HTTP message. HTTP defines the structure of these messages and how the client and server exchange the message (Edwards & Bramante, 2009).

Users browse the Internet using the WWW. The actual protocol that allows one to download images or other subjects from another web site is HTTP protocol, which uses the reliable service of TCP (Jain & Hassan, 2004).

HTTP has been the most popular internet protocol for 30 years. Until recently, its role has been limited to a traditional transfer of hypertext documents. However, its flexibility and interoperability cause it to be progressively involved in a much wider range of applications, from video and audio streaming to email, chat and documents editing (Augustin & Mellouk, 2011).

## **File Transfer Protocol (FTP)**

FTP is a file transfer protocol same as HTTP protocol, in fact, both have many common characteristics, for instance, they both run on top of TCP. However, the two application-layer protocols have some important differences. The most striking difference is that FTP uses two parallel TCP connections to transfer a file, a control connection and a data connection (Donahue, 2011). In other words, TCP based on IP networks include a file transfer application that allows users to send and receive

arbitrarily large files, these files may contain text, computer programs, images, or even digitized voice and video clips.

### **Simple Message Transfer Protocol (SMTP)**

It is at the heart of internet electronic mail, this protocol main job is to transfer's message from sender's mail servers to the recipient's mail servers. Currently, Microsoft's Outlook, Apple Mail, and Mozilla Thunderbird are the popular Graphical User Interface (GUI) agents for E-mail. Moreover, many text-based E-mail users interface in the public domain as well as Web-based interfaces E-mail (Kurose & Ross, 2010).

### **Real Time Streaming Protocol (RTSP)**

Allows a media player to control the transmission of the media stream, in more detail this protocol control actions include pause/resume, repositioning of playback, fast-forward, and rewind. The RTSP specification permits the messages to be sent over TCP or UDP protocol. After discussing some protocols that use TCP protocol, the next section describes the UDP protocol.

### **User Datagram Protocol (UDP)**

This protocol occupies the second rank in order of Internet usage after the TCP protocol (Kouvatsos, 2011). UDP is a no-frills, lightweight transport protocol, providing minimal services. "UDP is a best-effort-delivery protocol that sends data the best way it can, but doesn't take responsibility for the data's integrity. UDP is a store and forward protocol" (Curtis & Taylor, 2005). UDP is a connectionless

protocol, so there is no handshaking before the two processes start to communicate. This protocol “provides an unreliable data-transfer service-that is, when a process sends a message into a UDP socket, UDP provides no guarantee that the message will” ever reach the receiving process (Kurose & Ross, 2010).

Applications that do not require a dedicated stream to send data to a remote host often use UDP. UDP is faster, in that it does not require the overhead of establishing a connection between the two hosts, but it does not guarantee delivery of all data packets.

### **Domain Name System (DNS)**

This protocol is “name resolution service that translates the fully qualified domain name (FQDNs) into IP addresses. It consists of a system of hierarchical databases that are stored on separate DNS servers on all the networks that connect to the Internet” (Curtis & Taylor, 2005).

### **Simple Network Management Protocol (SNMP)**

This protocol is used to convey MIB (Management Information Base) information among managing entities and agents executing on behalf of managing entities. The most common usage of SNMP is in a request-response mode (Kurose & Ross, 2010).

## **2.3 Users Preference and Websites Categories**

A Web service is a software system identified by a URI, whose public interfaces and bindings are defined and described using XML. Its definition can be discovered by

other software systems. These systems may then interact with the Web service in a manner prescribed by its definition, using XML based messages conveyed by Internet protocols (Fan, Zhang, & Shen, 2010). As a web service develops and becomes more and more popular, many of them share the same or very similar function characteristics.

It is practical and universal to use a qualitative concept to describe a user's preference of QoS criteria. For instance, the reputation of web service is "important", and the execution price is "unimportant". But it is needed to convert the qualitative concept of quantitative value for further calculation (Fan et al., 2010).

On the other hand, the number of Internet users increases, they exchange and share information more and more frequently. In such an environment, reliable transactions are an important consideration. However, if individual users do not reflect their preferences, transactions cannot occur despite the user's trustworthiness. Considering this, most conventional research focuses on which providers are trustworthy (Gu, Yoo, & Park, 2008).

Web 2.0 is the future development trade of the Internet. Though it is new, it has been developing very fast, which has been proved by the China Internet Survey. Blog, social networking service (SNS), Podcast, are the most popular representative applications of web 2.0 websites, which are still immature (Dong, Clark, & Jacob, 2009).

As the Internet has grown, the roles have become blurred. For example, E-Commerce based on Web Service and Peer-to-Peer (P2P) networks are typical environments in which anyone can participate. In other words, many anonymous providers have appeared on the Internet in what is known as the Internet-based Overlay Network. In this network, users can construct their own logical networks without a centralized server, and here one of the critical issues is the selection of trustworthy providers (Gu, Hong, & Yoo, 2009).

### **2.3.1 Social Networks**

Social networks are gaining an increasing popularity on the Internet, with millions of registered users and an amount of exchanging the contents accounting for a large fraction of the Internet traffic. Due to this popularity, social networks are becoming a critical media for business and marketing, as testified by viral advertisement campaigns based on such networks (Canali, Casolari, & Lancellotti, 2010).

### **2.3.2 Blogs**

A blog is a type of website, maintained by an individual, or group with changeable entries of information, description of events, such as video or graphics. Entries are commonly displayed in a certain chronological manner (Oskouei, 2010). Furthermore, blogs considered as a new type of media that have recently become popular users on the World Wide Web and have influence throughout society. Developing an explanatory theoretical model of website usability is pivotal for understanding usable website design (Liao, Wang, & Tang, 2011)

### **2.3.3 E-commerce and Services**

The Internet has changed economic behavior at both the micro and macro levels (Xiaojian, 2009). Electronic commerce offers feasible alternatives to traditional ways of transacting business (Liu & Hu, 2009). Websites of firms are believed to be important to attract, convert visitors to, and retain customers. As a result, most firms have created and operated websites aligned with their business needs such as communicating with customers, promoting products, and processing orders (Fuentetaja & Economou, 2009).

The ubiquity of websites as a viable conduit of alternative transactions in business has led researchers to investigate factors that influence firms' adoption of e-commerce (Shan & Sun, 2011)

Firms provide various services via their websites based on their business strategies. This leads a firm's website to play a key role as a conduit to meet customers in an online space in a technology mediated service encounter (Xiaojian, 2009), in other words, as a proxy of a firm in an online space, a firm's website is competing against other websites provided by rival firms to attract and retain customers (Lee, Mirchandani, & Zhang, 2010).

## **2.4 Related Work**

The importance of network measurement lies in the observation and understanding of networks. There are various studies focusing on this topic, both in passive and active modes of measurement. Some of the focus is concentrated on measuring specific kinds of protocol as the one conducted by Yang, Chen, & Jin (2011) which

is dedicated to high speed real-time passive HTTP traffic performance measurement. They brought forward a real-time architecture to measure HTTP passive performance and considered benchmarking to ensure future work's accuracy. On the other hand, some other studies concentrated on classification of Internet traffic (Kim, Claffy, & Fomenkov, 2009) In study, critically evaluating traffic classification, Kim, Claffy, & Fomenkov (2009) conducted a thorough evaluation of three classification methods on the basis of transport layer ports, host behavior and flow features. The aim of the study is to highlight the effectiveness of port-based classification through the identification of the legacy applications supported by the utilization of packet size and TCP flag information.

While other studies were focused on measuring application's behavior over Internet traffic (Iliofotou, 2009) This study proposed a graph-based representation of internet traffic which captures the network wide interactions of applications. They worked on TDGs in depth and they provide a graph that summarizes network-wide behavior of applications in order to classify network traffic .

Other studies were directed to P2P applications (Cao et al., 2010) in this study authors focused on analyze three locality-awareness policies for BitTorrent-like system: tracker locality, choker locality, and picker locality. Based on analyze how much network load saving can be expected for these locality policies, as well as their impact to the downloading efficiency of the system, as a result they become with that all locality policies can significantly reduce the average distance of both downloading and streaming scenarios.

Another study (Wang, An, & Yang, 2010), authors study internet traffic on the Tsinghua University campus network they presented characteristics of inbound traffic flows from the aspects of traffic prediction and inference. Then they analyzed the geographical origins of incoming flows, and the result reveals that the USA, Japan and Korea are the most important source countries of international traffic, they investigated on the external traffic of a large campus network in China based on a one-year-long flow-based data set and a 10-year-long user-based data set.

They characterized the traffic pattern of four incoming traffic groups, furthermore they demonstrated that traffic prediction of domestic flows based on historical data is much reliable than international flows, and the number of active hosts in the campus network is more likely to resemble the past than other statistics. They also displayed that the linear relationship does not hold for most statistic pairs, and only the relationship between traffic rate or flow number and packet number exhibits a nearly linear property for all traffic groups.

In result they also find that 74 percent of the international traffic volume is coming from only 5 countries, where in USA ranks first. There are only 7.3 percent of international http active source hosts in Japan, however it provides 27.2 percent of http traffic, and results show that the major has a stronger influence of users' average online time, while occupation has a stronger influence on users' average international traffic volume.

Another study (Augustin & Mellouk, 2011) authors studied 20 popular, Web applications that are representative of 12 application types. They describe a method



to isolate and capture browser-generated traffic and plot time series with an RRDTool database.

As a result of this study authors proposes a preliminary draft for a classification of Web applications according to the three traffic features traffic intensity is the total volume of traffic exchanged by the application during the measurement period, traffic symmetry measures the ratio between upstream and downstream traffic and traffic shape describes the general aspect of the traffic pattern. However, the work presented in this paper is only a first step towards a precise classification of the plethora of HTTP applications available on today's Web.

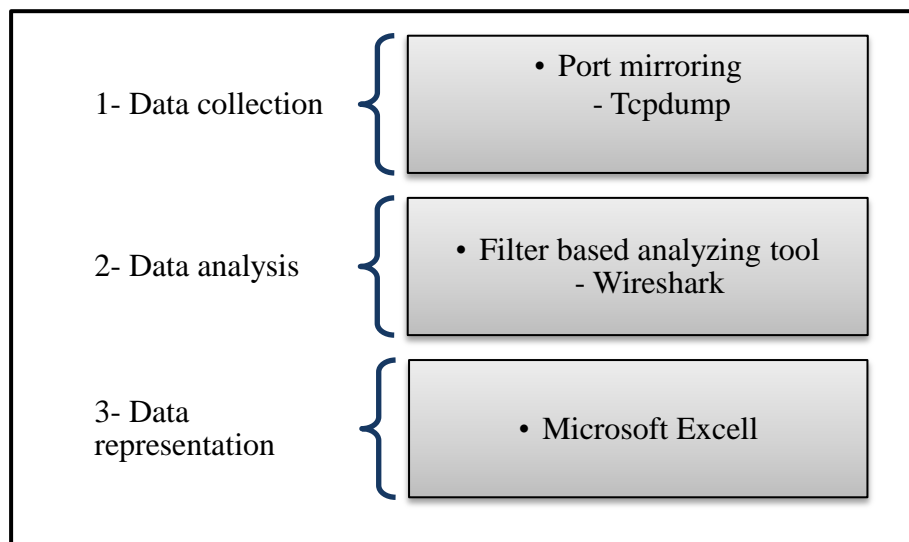
## CHAPTER THREE

### PROJECT METHODOLOGY

#### 3.1 Introduction

The chapter explains the measurement tasks utilized to collect data for valuable information which will help achieve the study objectives and provide an effective solution to the problem.

Section 3.2 elaborates on the network data collected and UUM network topology and devices utilized, while section 3.3 provides a description of the method and the tools that are analyzed. Finally, section 3.4 provides a description of the software utilized to present the results.



*Figure 3.1: Project Methodology (Jain & Hassan, 2004)*

The phase is characterized by the collection of raw data from the operational network where the type and amount of data collected hinges on its use (Jain & Hassan, 2004). Prior to initiating data collection process, there are two main aspects that have to be taken into consideration – network topology and devices. It is imperative to determine the most suitable place for collection of packets in order to avoid capturing irrelevant packets.

### UUM – Overview Schematic Design

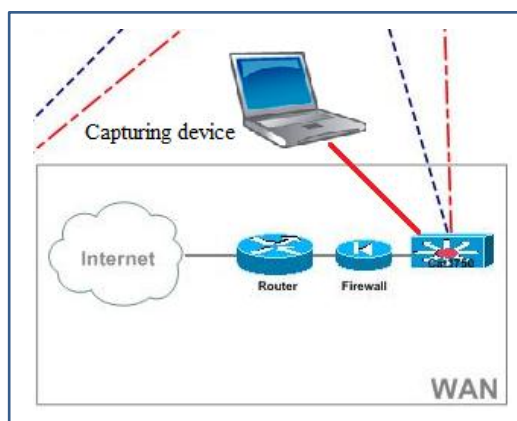
The diagram illustrates the UUM network architecture, showing a central core (Pusat Komputer) connected to various campus networks and a WAN. The legend defines the components and line styles used:

- Switches:**
  - Core Switch (Blue star icon)
  - Dist. Type I Switch (Blue star icon)
  - Dist. Type II Switch (Blue star icon)
  - Core Switch (Blue star icon)
  - User Node(s) (Blue square icon)
- Line Styles:**
  - 1000 BASE-LH (SMF) (Blue dashed line)
  - 1000 BASE-SX (MMF) (Red dashed line)
  - 10/100/1000 BASE-TX (STP/UTP) (Black solid line)
  - 10/100 BASE-TX (UTP) (Black solid line)
  - 10000 BASE-LR (SMF) (Blue dashed line)
  - 10000 BASE-SR (SMF) (Red dashed line)

28

Lenovo G560 was used as a capturing device and was connected to the main distribution multilayer switch in UUM computer center, the type of this switch is Cisco's catalyst 1750, which supports mirror porting. This switch is connected to two main multilayer switch Cisco's catalysts 6509 by fiber-optic data-transfer cables.

The main multilayer switches are connected with all distribution's switches across the university divisions, including two distribution switches that are connected to the servers who operate in the network; furthermore, it connected to three controllers' devices. Most of the distribution multilayer switches are Cisco's catalyst 3750 which in turn are connected to layer two Cisco's catalyst 2960 switches. Students who use wire connection inside UUM campus are connected to these layer two Cisco switches, unlike the student who uses wireless connections. They are connected to distribution multilayer catalyst 3750 switches. Figure 3.3 refers to the location of the capturing device.

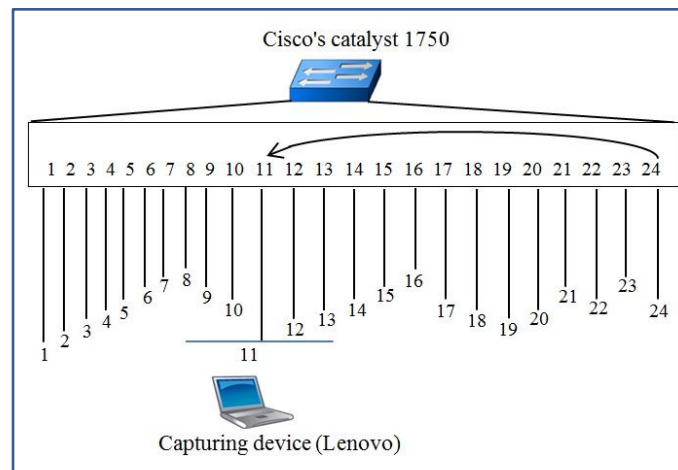


*Figure 3.3: Capturing Device*

Data have been collected from UUM main multilayer switch (Cisco's catalyst 1750), Port mirroring used to capture the data. Mainly, Port mirroring used to direct a copy

of network packets from single or multiple ports to another switch port, whether the packets were normal network packets or VLAN (Team, 2006). Port mirroring works with routers and switches; it can be remotely or directly. The most recent routers support this feature. Routers send the mirrored packets directly to the destination port or may transform it to the internal memory then resend it. This situation occurs in case there was a high load of network packets on the mirrored port.

In UUM distribution switch Cisco 1750, one particular switch interface Gigabit-Ethernet 1/0/24 was mirrored, this port connected to the bandwidth manager which is connected to the firewall in the computer center. This interface Gigabit-Ethernet 1/0/24 mirrored to Gigabit-Ethernet 1/0/11 interface, capturing packets occurred from this interface. The port that has been mirrored and destination port displayed in Figure 3.4



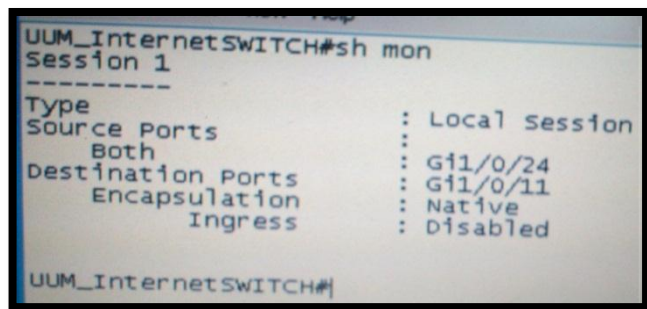
*Figure 3.4: Port Mirroring*

Previous Figure 3.4 displays all working ports within the switch, including the mirrored port (24) and the destination port (11) which capturing device was connected to. The following Figure 3.5 shows the mirroring process codes:

```

UUM_InternetSWITCH
(config)#monitor destination 1
destination interface gigabitEthernet
1/0/11
UUM_InternetSWITCH (config)#end
UUM_ UUM_InternetSWITCH #sh
mon
Session 1
-----
Type                : Local Session
Source Ports        :
  Both              : Gi1/0/24
Destination Ports   : Gi1/0/11
  Encapsulation     : Native
  Ingress            : Disabled
UUM_InternetSWITCH#

```



```

UUM_InternetSWITCH#sh mon
Session 1
-----
Type                : Local Session
Source Ports        :
  Both              : Gi1/0/24
Destination Ports   : Gi1/0/11
  Encapsulation     : Native
  Ingress            : Disabled
UUM_InternetSWITCH#

```

*Figure 3.5: Port Mirroring Codes*

### 3.2.2 Packets Capture

Linux provides a unique and conducive environment to both packet capture and software development given its open source nature. This packet analyzer takes advantage of both the Linux environment and its open source software (Crandall & Jasani, 2011).

Capturing device use Linux Ubuntu 12.04 LTS 64 Bit operating system and Tcpdump tool to capture mirrored packets, tcpdump running on the capturing laptop to capture all network traffic from specific interface using this command:

- `sudo tcpdump -i <interface> -C <file-Size> -w <file name and path>`
- Ifconfig used to view all interfaces and other information.
- This command was used to capture the data.

```
sudo tcpdump -i eth0 -C1000 -w /home/mustafamh/Desktop/CFOR/CAPTURE/1.cap
```

While `-i` to determine which interface used to capture the data, `-C` to generate 1 GB files size and `-w` to specify where to store the captured packets.

Data have been captured from the network throughout all working days for a full week, starting from Sunday until Thursday, for one hour from ten o'clock until eleven in the morning. Total file size of captured packets were 197.2 Giga-Byte divided as follows:

*Table 3.1: Size of Captured Data*

Day	Date	Total Files Size	From To
Sunday	15/4/2012	37.8 Giga-Byte	10:00 until 11:00 AM
Monday	16/4/2012	36.3 Giga-Byte	10:00 until 11:00 AM
Tuesday	17/4/2012	41 Giga-Byte	10:00 until 11:00 AM
Wednesday	18/4/2012	39.1 Giga-Byte	10:00 until 11:00 AM
Thursday	19/4/2012	43 Giga-Byte	10:00 until 11:00 AM

After collecting data for each day, packets were sorted and isolated by timestamp, any packets carry timestamp before ten o'clock exactly has been ignored as are the packets that have been taken after eleven, which was the reason of the fraction number in Table 3.1.

### **3.3 Data Analyzing**

All data collected from the network is raw data divided and saved in the form of (Cap) extension files, libpcap library was developed using C programming language designed to convert network interface card NIC in promiscuous mode, that's guaranteed that all packets arrived to the interface will be captured.

### **3.3.1 Wireshark Program**

Wireshark is the most well-known open-source network data analyzer available. Wireshark is a steady and valuable component for all network toolkits (Orebaugh et al., 2007). Wireshark was selected for the packets analyzing process because of:

- Open source program.
- Works on different operating systems (Linux, windows, Mac).
- Compatible with Tcpdump files (The ability to read all output files) .
- Allows scripting and plugins.
- Support and analyze more than 850 different protocols.

### **3.3.2 Using Wireshark to Achieve the First Objective**

The Wireshark is utilized to achieve the first objective. Hence, 38 files were collected from the internal network of the UUM, with the size of each file being 1GB except the last one (804MB). The analysis was conducted to determine the number of packets for every file and to measure the speed and size of each packet contained in each. The average is taken by dividing the statistics on the number of files. The analysis processes were similar throughout the five days; with the differences lying only in the number of packets, size of packets and the number of files.

Frame length analysis: frame.len filter was used to analyze the frame length for each packet, depending on the frame length packets were classified into groups, whether 40-79 or 80-159 or other lengths.



Internet protocol analyzer: IPv6 and IP filters were used to isolate all IPv4 and IPv6 packets. Moreover, it used to calculate the packet's number and size from both types. As described in chapter four, Internet protocol version four.

This protocol was analyzed from different aspects: Calculating the packet's and size of open shortest path first protocol (Routing Protocol) by using OSPF filter, furthermore, calculating the packet's number and size of the address resolution protocol (ARP) by using ARP filter as well as calculating the packet's number and size of generic routing encapsulation protocol (GREP) by using GRE filter, moreover, analyzing another protocol such as TCP, UDP, ICMP, HTTP.

The following Table 3.2 summarizes the most important filters used in the process of analyzing the packets and its function:

*Table 3.2: Wireshark Filters (Orebaugh et al., 2007).*

Filters	Functions
IP and IPv6	To isolate all IPv4 packets from IPv6 Packets.
frame.number	To calculate packets number of each (.cap) file.
frame.len	To calculate packets length.
frame.time	To sort the packets according to time (second).
ip.addr == 10.0.0.115 and ip.addr==85.12.10.2	To get all conversation packets between two hosts, whether the conversation occurred by hosts using IPv4 or IPv6.
OSPF	To specify and isolate all OSPF Protocol packets.
ARP	To specify and isolate all ARP Protocol packets.
GER	To specify and isolate all GERP Protocol packets.
TCP	To specify and isolate all TCP Protocol packets so other protocol that fall under this protocol can be analyze later like HTTP and RTSP...etc.
UDP	To specify and isolate all UDP Protocol packets so other protocol that fall under this protocol can be analyze later like DNS and EDonky...etc.
ICMP	To specify and isolate all GERP Protocol packets.

### 3.3.3 Using Wireshark to Achieve the Second Objective

Each packets file for each day analyzed separately by following these steps:

- Open each libcap packets file.
- Isolate HTTP protocol and save all HTTP packets in different files.
- Isolate HTTP request from an HTTP by using Http.request filter and save.
- Calculate websites statistic from each HTTP request packets file by using (IP contains domain-name) filter.
- Isolate each domain-name packets and save it to libcap file.

Table 3.3 summarizes the most important filters used in the process of analyzing Http request packets and its function:

*Table 3.3: HTTP Filters (Orebaugh et al., 2007).*

Filters	Functions
IP	To specify and isolate all IPv4 packets
HTTP	To specify and isolate all HTTP protocol Packets for each day.
Http.request	To isolate all HTTP requests from all http packets
IP contains domain-name	To specify all IP headers that contain particular domain name.

After using all filters on each HTTP request file, which contained the packets for each domain name, statistics such as packet's number and average size were collected and placed on Microsoft excel tables for later analyze. Figure 3.6 illustrates the use of (IP contains Google) filter on Wireshark open-source program:

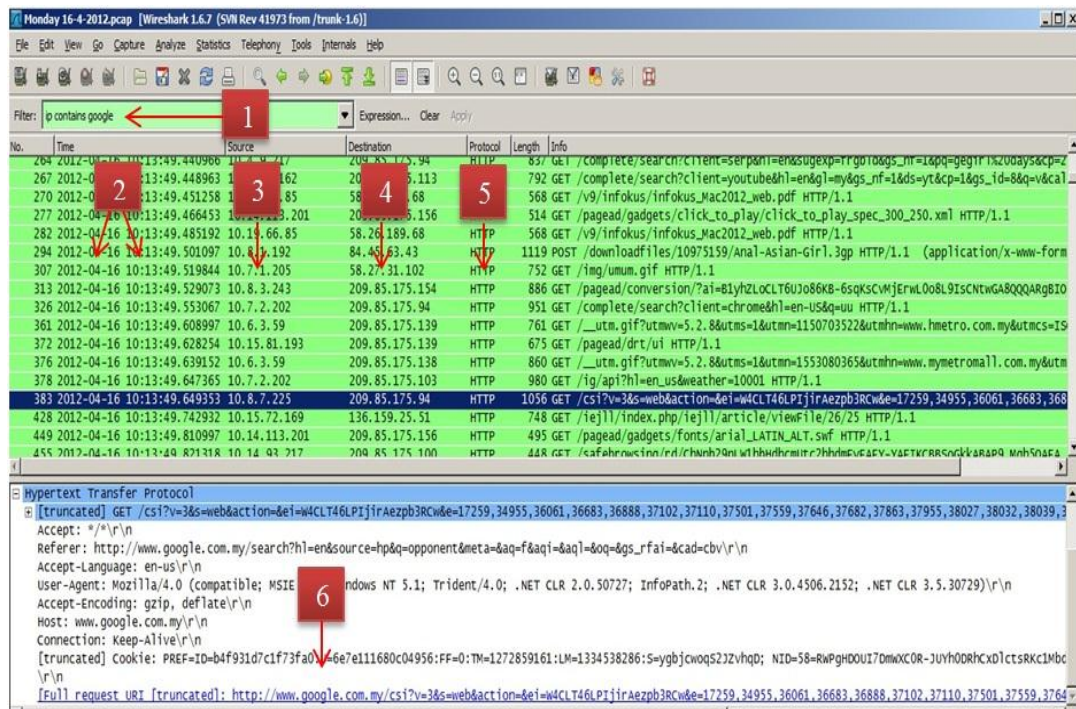


Figure 3.6: Wireshark, IP Filters

Previous Figure briefly describes the packets filtering process using Wireshark program where the numbers refer to:

- 1) Filter textbox where IP contains Google filter used.
- 2) The date and time for each packet, marked packet was taken on Monday 16/4/2012 at 10:13:49.65 AM.
- 3) Source IP address, marked packet was sent from 10.8.7.225 IP address.
- 4) Destination IP address, destination IP for marked packet was 209.85.175.94.
- 5) Type of protocols, in this Figure refers to the HTTP protocol.
- 6) Refer to the domain name (Google).

After completing these steps all Google domain-name packets were isolated, and sub-domain filters were used to specify all Google sub-domains such as mail.google.com and other sub-domains.

GeoIP Domain Name and GeoLiteCountry databases were implemented in the Wireshark program to specify each country packets and domain name by following these steps:

- 1- Download GeoLite and GeoIP (purchase) from MaxMind.com
- 2- Wireshark version 1.6.2 was used, previous databases support this version.
- 3- The name resolution location was changed in Wairshark.
- 4- This filter was used to gain a specific country packets:

ip.geoip.country == "country name"

A similar process in Figure 3.6 was used to achieve the first objective by changing filters. The next section describes the data representation software:

### **3.4 Data Representation**

It produces graphs and charts to present performance metrics for visual use (Jain & Hassan, 2004). Microsoft Excel program used to presents the data. This program can deal with mathematical problems, tables and presents graphical Figures.

Using Microsoft Excel to presents first and second objective: main step was transferred all packet statistics, that's taken from different stages of analyzing process to Excel spreadsheets, rows and columns have specified in ways that allow

drawing Figures, which exactly reflect the packet's status. Following Figure displays the statistics of OSPF protocol table in Excel spreadsheets:

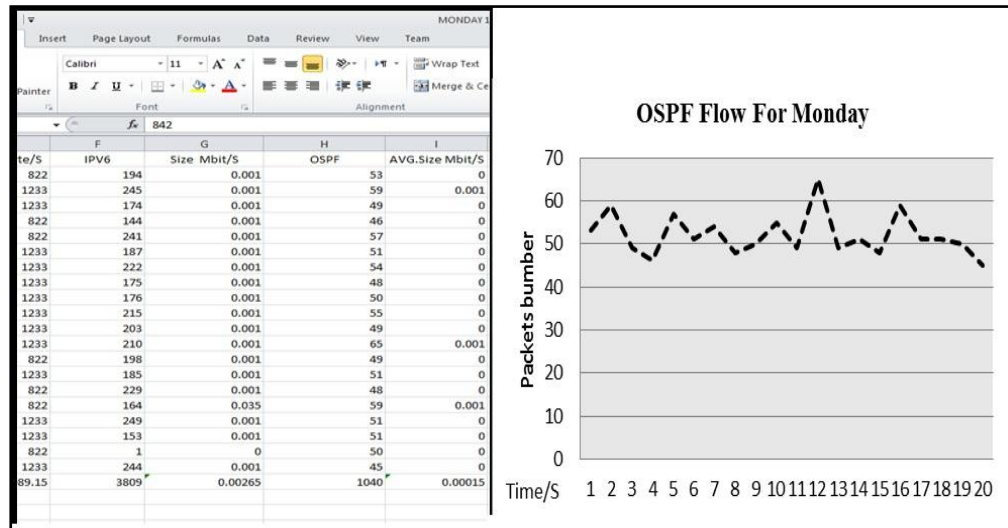


Figure 3.6 OSPF Flow on Monday

Figure 3.6 describes how statistics can be used to create a Figure that presents the packet's status, on the right side of the Figure; two variables were used, the number of packets and time per second. These variables were taken from the table shown in the left side of the Figures. Previous method was used to represent all domains' packets, country's packet. Moreover, to illustrate users' preferences and to measure network performance, all packets were presented by using this method, for both first and second objectives of this study.

## **CHAPTER FOUR**

### **NETWORK PERFORMANCE**

#### **4.1 Introduction**

The present chapter explains the achievement of the first objective mentioned in chapter one which is measuring the network performance of UUM campus to determine which day took the largest number of packets, the distribution of network protocols and its uses. The following section (Section 4.2) describes the variables measured in the UUM network while section 4.3 provides a description and analysis of different kinds of protocols in an attempt to specify the protocol that took the greatest number of packets. Finally in section 4.4, the chapter summary is provided.

#### **4.2 Network Performance**

Table 4.1 presents the numbers and the average size of packets through one hour for each day:

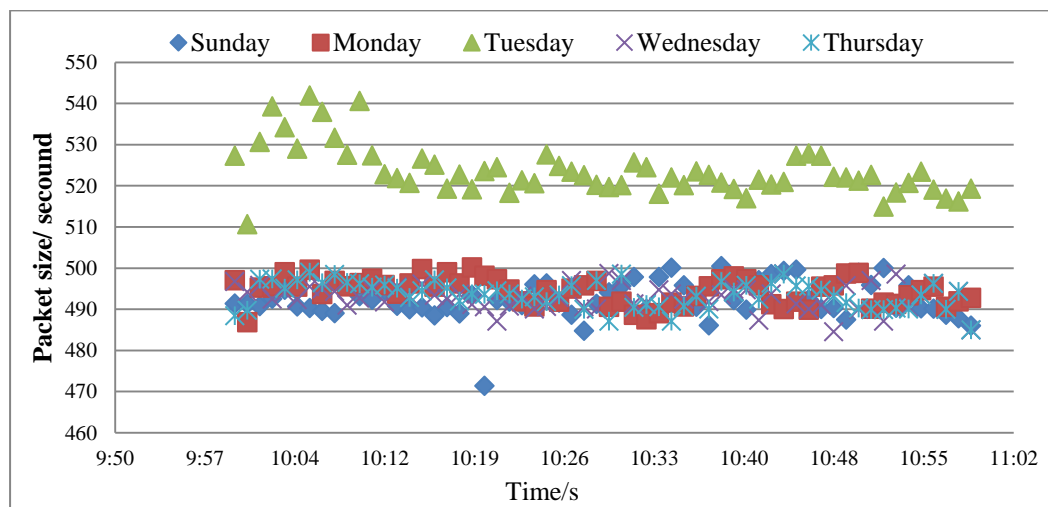
*Table 4.1 All Day's Statistic*

Statistic	Sunday	Monday	Tuesday	Wednesday	Thursday
Packets Number	78805385	78657343	83164352	79535383.99	85430093
Avg Packet size per second	492.2092089 K-byte	494.385683 K-byte	523.6448711 K-byte	492.837357 K-byte	493.52574 K-byte
Avg packets per second	21890	21849	23101	22093	23730

From Table 4.1 Tuesday and Thursday had the highest number of packets, while in the other days Sunday, Monday and Wednesday the packets number was very convergent. Knowing the two parameter packets number and average packet size per time, make possible to measure the throughput of the network (Jain & Hassan, 2004).

#### 4.2.1 Network Load and Throughput Measuring

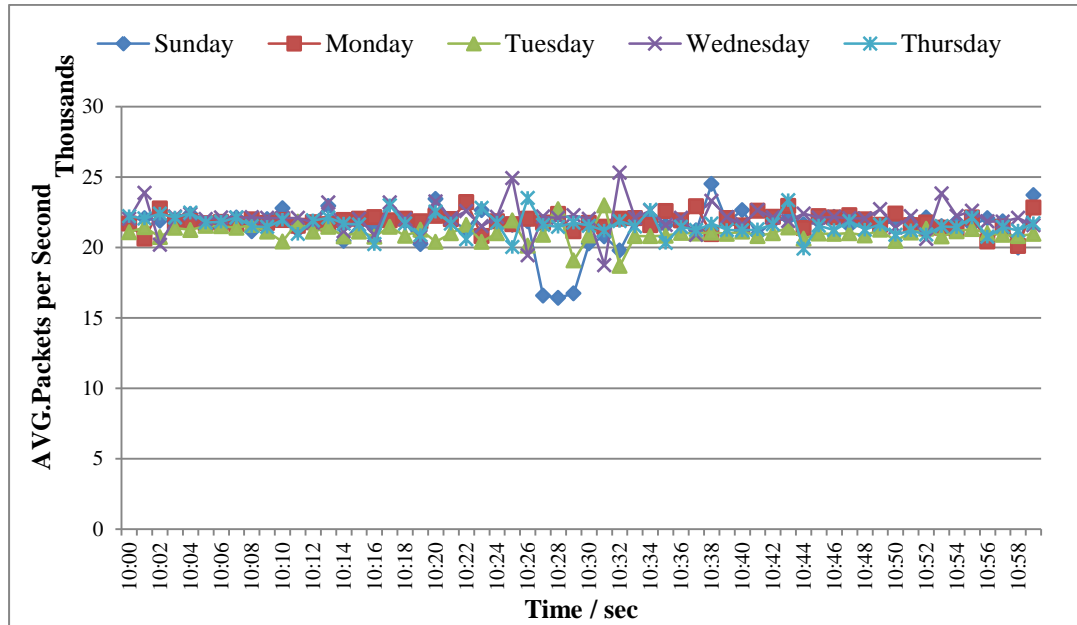
Most typically method to measure network throughput is measuring the average packet size per second (bit/s or bps). Data packets per second (P/S) network load, network load refers to the amount of traffic being carried by the network (Kim, Claffy, & Fomenkov, 2009). Figure 4.1 presents the throughput of all days.



*Figure 4.1: Throughput of All Days*

From Figure 4.1, it is noticeable that high size of packets transmitted within hour on Tuesday from the rest days of the week, the reason was because of the high usage of some protocols (or software that uses these protocols), HTTP protocol for an

example has the highest rate of use on Tuesday, used rate for each day for the rest of protocols, will be clarified throughout this chapter.



*Figure 4.2: Network Load within Hour of Each Day*

The previous Figure 4.2 shows the network load per one hour of each day, traffic on Sunday was normal except during the time period 10:26 to 10:29 AM the traffic was lower than usual, this due to one or more network devices become temporarily out of service, so the main switch didn't receive traffic from this device, on Monday traffic flow was usual except in one second recorded at 10:59:13 AM Average packets was 23165. On Tuesday traffic flow was normal until 10:29:40 AM low traffic was recorded and at 10:31:53 AM also. On Wednesday, at 10:01:54 high traffic was recorded as well as at 10:23:41 and 10:30:17AM. On Thursday traffic flow was as usual. The next section describes packets loss in the campus network.



#### 4.2.2 Packets Loss

Packets loss caused by transmission errors is a very common issue. Table 4.2 illustrates packets loss throughout one hour of each day:

*Table 4.2 Packets Loss*

Day	Packets loss	Packets per one hour	Packets loss rate (%)
Sunday	226800	78805385	0.002877976
Monday	448980	78657343	0.005708049
Tuesday	558900	83164352	0.006720428
Wednesday	412953	79535383.99	0.005192066
Thursday	625193	85430093	0.007318182

There are various reasons for packets losses during transmission, the major source of packets loss are 1- buffer overflow at the intermediate routers, and 2- packet corruption caused by transmission errors (Jain & Hassan, 2004). The process of analyzing the loss packets shows that the main corrupted packets were Secure Sockets Layer (SSL), Graphics Interchange Format (GIF), Domain Name Service (DNS) and Hypertext Transfer Protocol (HTTP) protocols packets.

#### 4.2.3 Packets length distribution

Packet length distribution classifying the rate and percentage of each packet length group. Packets length characterizing is valuable in network troubleshooting, when large amounts of small packets can place extra liability, load and tasks on networking equipment that leads to reduced throughput (Orebaugh et al., 2007). Table 4.3 presents packets length:

*Table 4.3 Packets Length*

Length /Days	Sunday	Monday	Tuesday	Wednesday	Thursday
40-79	40571814	43166340	42908219	61459730	44618183
80-159	7208277	5740972	7129754	6710252	7485389
160-319	1941744	1713894	1922975	2121599	1958226
320-639	3648335	3318776	3363878	3835739	3729659
640-1279	8898391	7932630	9790381	8265392	10049655
1280-2559	16391204	16784029	18048595	16269686	17534664
2560-5119	617	698	550	592	314
5120+	3	4	0	0	3

Maximum packets length for IPv4 over Ethernet network is 1526 bytes (1500 byte payload + 14 byte header + 8 byte preamble + 4 Byte cyclic redundancy check). Maximum transmission units specify IP packets that can be transmitted without fragmentation. IPv4 packets should be greater than or equal to 68 bytes to transmit. Furthermore, IPv6 packets took at least 1280 bytes to transmit, Ethernet jumbo packets took between 1500-9000 byte (Mieghem, 2009). Previous Table displays the number of packets that were addressed by the network devices. Most of these packets were short-length 1GB from collected data were analyzed due to this process most short-length were HTTP, ICMP, and SS Layer packets, nevertheless, UDP Protocol took most of the long-length packets. Figure 4.3 displays packets length for each day:

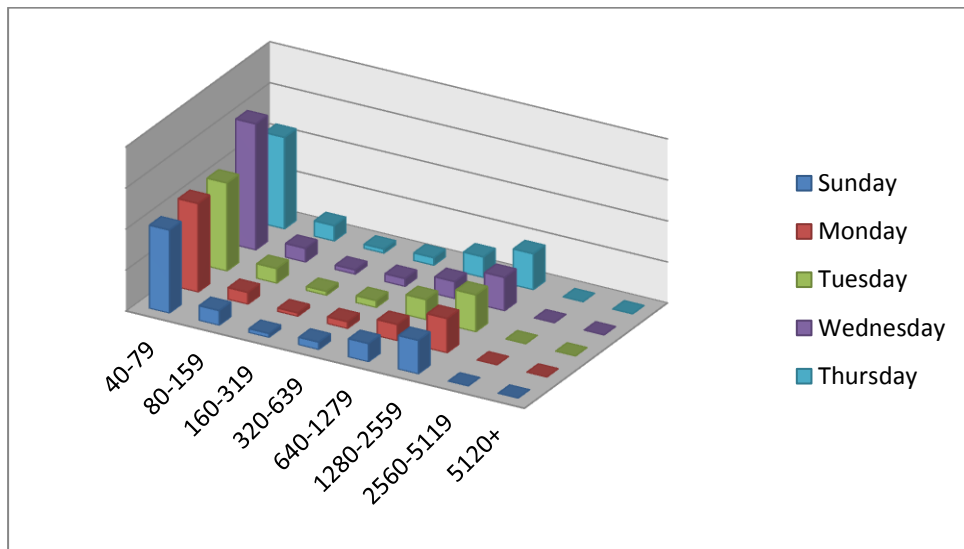


Figure 4.3: Packets Length for All Days

### 4.3 Protocols Distribution

This section defines most important protocols that were operate in UUM network in order to summarize which protocol acquired highest packets rate and how much each protocol reserved from network bandwidth. The analysis process is handled through several stages illustrated in the following Figure:

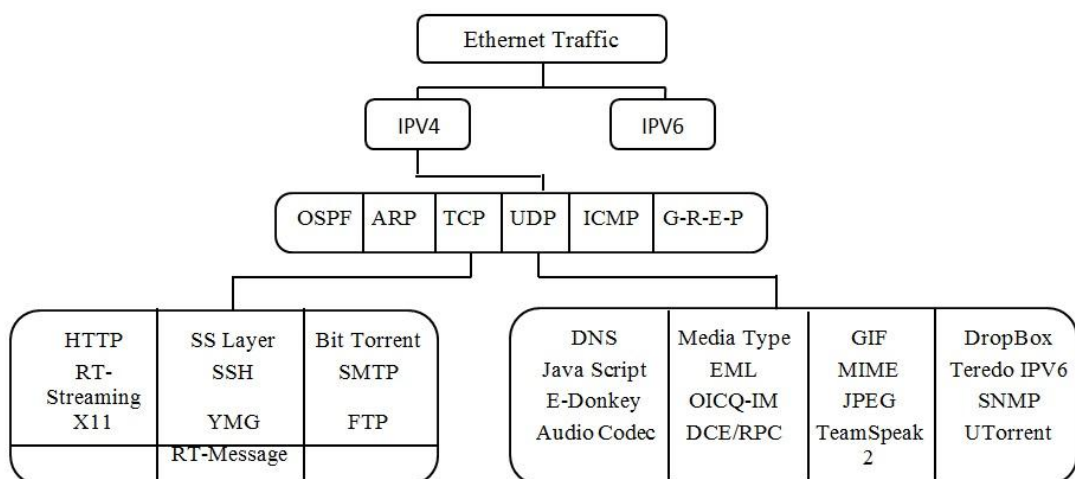
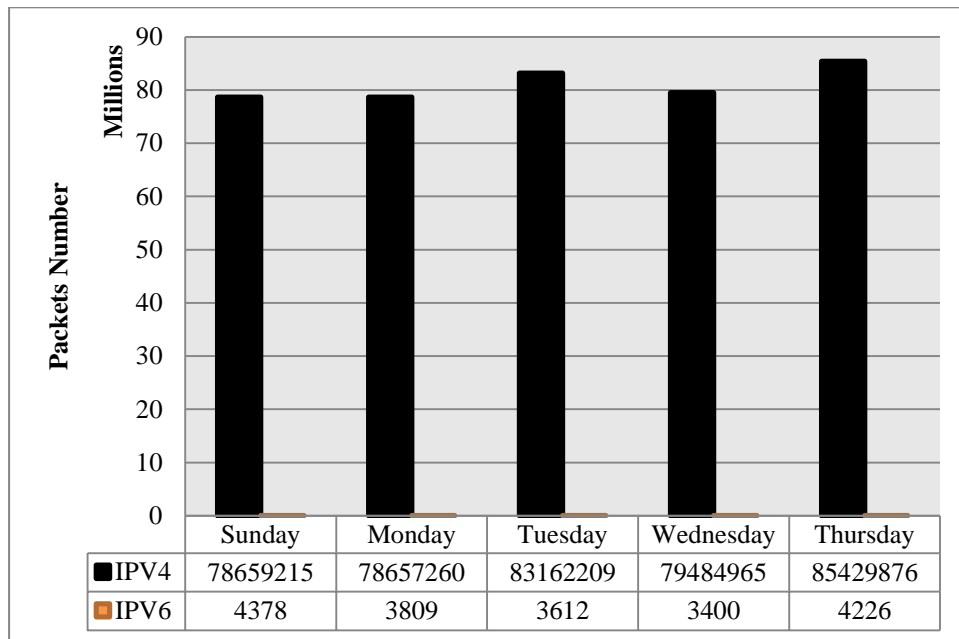


Figure 4.4: Protocols Analysis Stages

Previous Figure specify each step analyzing protocols stages, at first it starts with Ethernet traffic in general, afterward it progressed to the IPv4 analyzing stage, in the meantime all packets that's related to each protocol was isolated, finally, the process of analyzing reach the last stage which is TCP and UDP protocol analyzing.

#### 4.3.1 Ethernet Traffic Distribution



*Figure 4.5: Ethernet Traffic Distribution*

Through this process of studying all Ethernet packets which have been taken from the internal network of the University, main protocols were summarized to identify which protocol took highest packets and the size of the bandwidth, the previous Figure reviews the results. IPv4 gained the highest packet number on the contrary with IPv6 packets. This protocol will be studied later in detail to define what kind of protocols that operates under IPv4 protocol.

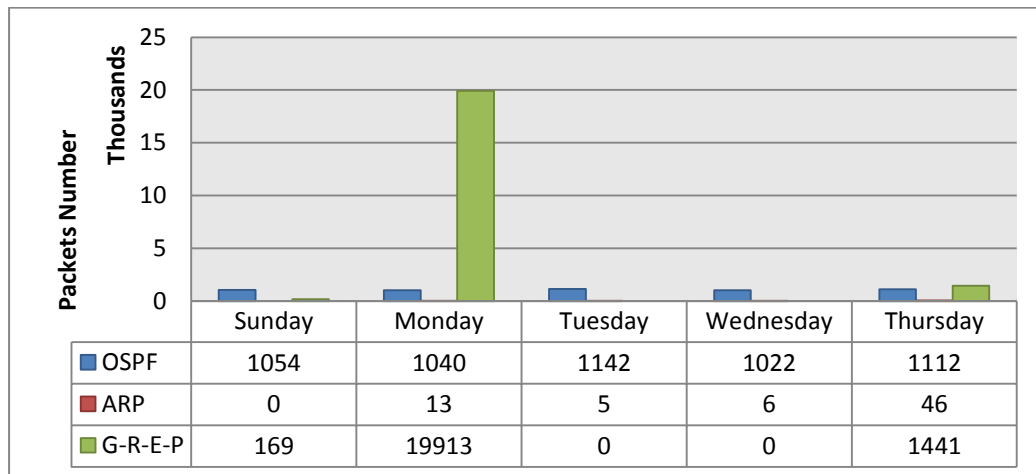
*Table 4.4 Ethernet Average Packets Size*

Statistics	Sunday	Monday	Tuesday	Wednesday	Thursday
IPv4	86.22	87.23	92.25	86.76	87.47
	M-Byte	M-Byte	M-Byte	M-Byte	M-Byte
IPv6	0.000952381	0.00265	0.001	0.001	0.001
	M-Byte	M-Byte	M-Byte	M-Byte	M-Byte

Table 4.4 shows the acquisition sizes of each protocol on the network bandwidth, ratios were calculated for each protocol during full hour from 10:00 AM to 11:00 Am from each day. The average percentage achieved by the IPv4 protocol was 92.7 MB for one hour on Tuesday; in fact UUM Internet bandwidth is 160 MB and Ethernet highest traffic size rate was 92.721 MB, therefore it is reasonable to say that the Internet bandwidth can accommodate daily use according to these statistics, on the other hand, average rating and capturing time limitation has to be taken into account.

#### **4.3.2 IPv4 Distribution**

Through the analyzing process, all IPv4 Packets were isolated from IPv6 packets in order to facilitate the process of analyzing and to reach more accurate results. After classifying the rates of IPv4 protocol, this protocol still the most broadly deployed Internet protocol (Dulaney & Harwood, 2012). During this phase IPv4 protocol has been studied extensively to demonstrate the most important protocols that fall under it and determine the usage percentage of each protocol, following Figure illustrates most important protocols



*Figure 4.6: OSPF, ARP and GREP Packets Numbers*

UUM distributed multilayer switch connects to the rest of the switches in the local LAN in order to make the process of exploration of the network in the campus. Thus, each multi-layer switch gets routing-table updates to complete the data transmission. Total send and received OSPF packets during all days were few compared to the total number of packets sent through each day. Therefore, its effects will be slight or may be absent on the network switches. All OSPF average packets size for all days during one hour was 597.48 KB per second, each OSPF data were in byte value when it was collected, and then it was converted to KB by dividing data into 1024. After analyzing the packages it became clear that most packets fall under Hello protocol and carry 224.0.0.5 IP Address.

### **Address Resolution Protocol (ARP)**

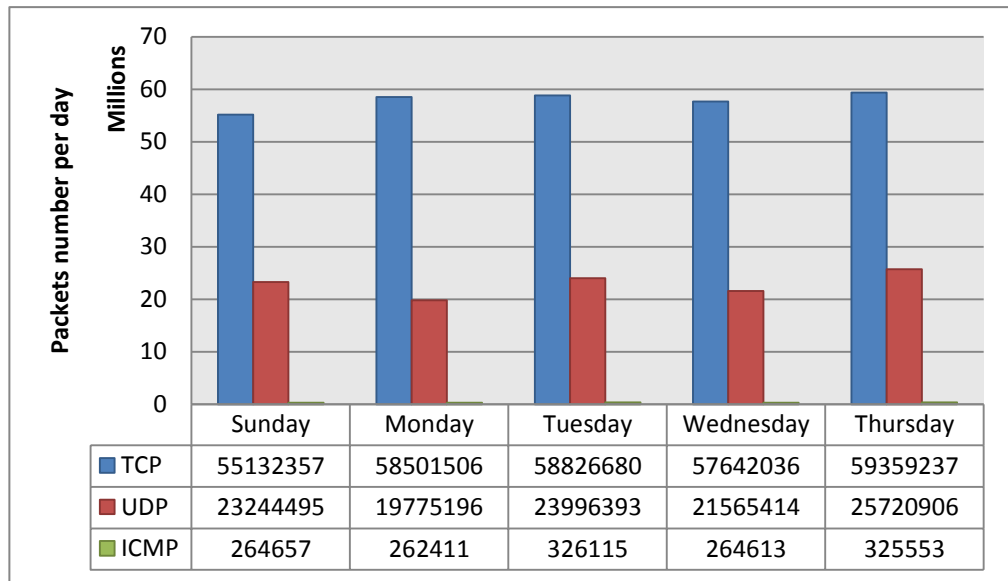
Main task for this protocol is to translate an IP address to a Media Access Control (MAC) for any host that is on the same broadcast local-area network (or, subnet) (Marsic, 2010). The previous Figure displayed the statistic for ARP protocol, there are no transferred packets on Sunday, and the highest number of packets was noted

on Thursday, where 46 packets only were registered. The number of data considers slight when compared with the total number of Packets per day, because the multi-layer switch doesn't interfere with all ARP communications happens between Hosts and Switches.

### **Generic Routing Encapsulation Protocol**

The main function for GREP is creating a tunnel it used when packets has to be sent through one network to another, without being analyzed or evaluated similar to IP packets by any intervening network equipment. The GREP protocol does not encrypt the data, for this reason it usually conjunction with point to point protocol (PPP) to create and exchange virtual private network (VPN) information (Donahue, 2011). This is based on the principle of linking different networks or hosts remotely through the Internet. Furthermore, GRE protocol is used within IPsec protocol, which is a protocol used to protect IP protocol data by providing identity verification and encryption for each packet commonly it's used to permit passing of routing packets amongst networks. Also this protocol used by some operating systems like Linux to establish ad-hoc IP over GRE which can operate with Cisco devices.

The previous Figure illustrates that Monday got the highest GREP data transmission this is because one of the three factors that had been mentioned previously, it took 0.137 percent from the total packets exchanged that day. After analyzing the captured data, most GREP packets were used for PPTP (point to point tunneling protocol) this mean that a VPN was created and start exchanging information.



*Figure 4.7: TCP, UDP and ICMP Packets Number*

Transmission control protocol is one of the most essential protocols that work in the Internet field. Ordinary any Internet network obtains a high percentage of TCP packets because this protocol is highly used by Internet applications such as E-mail, World Wide Web and file transfer applications (Dulaney & Harwood, 2012), however from the previous Figure TCP gained the highest packet rates from other protocol, especially on Thursday and Tuesday.

Afterward UDP protocol took the second rank, all applications that do not necessitate reliable data stream may use this protocol such as domain name system, voice over IP and online games, Thursday and Tuesday gained the highest of UDP packets rates, on the other hand ICMP protocol took the third rank for the same days also. TCP and UDP protocols will be analyzed in detail later in this chapter. Table 4.5 shows the usage rates of each protocol in the network:



*Table 4.5 IPv4 Average Packets Size*

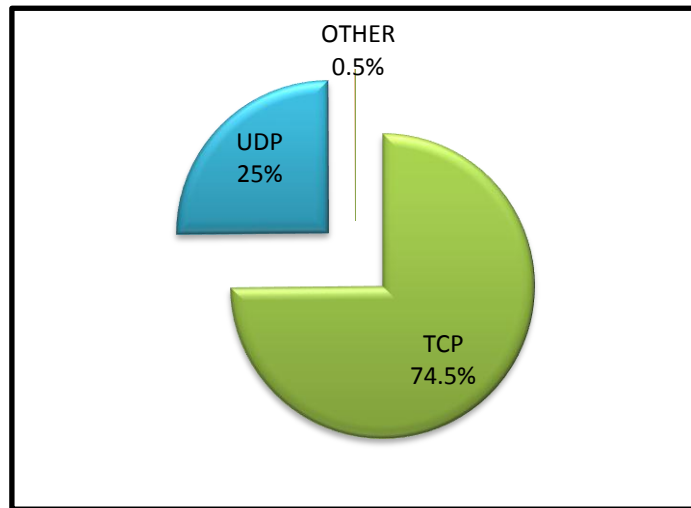
Day Protocol	Sunday	Monday	Tuesday	Wednesday	Thursday
OSPF	0.000142857 M-Byte	0.000165 M-Byte	0.00014 M-Byte	0.00015 M-Byte	0.000212727 M-Byte
ARP	0	0	0	0	0
G-R-E-P	0.000531 M-Byte	0.0425 M-Byte	0	0	0.00064 M-Byte
TCP	64.50995238 M-Byte	68.697605 M-Byte	70.69177273 M-Byte	66.4123 M-Byte	64.85663636 M-Byte
UDP	21.62185714 M-Byte	18.45215 M-Byte	21.94781818 M-Byte	20.2999 M-Byte	22.05577273 M-Byte
ICMP	0.052285714 M-Byte	0.054 M-Byte	0.060636364 M-Byte	0.05085 M-Byte	0.061 M-Byte

Table 4.5 demonstrates the average packets size for each protocol during the capture time for all working days, OSPF packets took the lowest rates, however this is a normal situation because routing protocol packet sizes are very slight (Hartpence, 2011), besides of that from Figure 4.6 all OSPF packets does not exceed 1112 packets per one hour, because of these facts this protocol has no or limited effects on the network.

ARP has a very slight number of the packets, however packets size didn't reach 20 KB. The size of GREP packets has reached in the highest point 43.52 KB per on hour on Monday, possible reason behind that is creating and exchange VPN packets through UUM Network .

Heights TCP average packets size was 70.7 MB per one hour on Monday, TCP protocol carried thousands of packets per one second, possible reasons for this high

usage will discuss later in TCP analyzing section, furthermore UDP highest average packets size was 22 MB per one hour on Thursday. Following Figure describes the average size for each protocol:



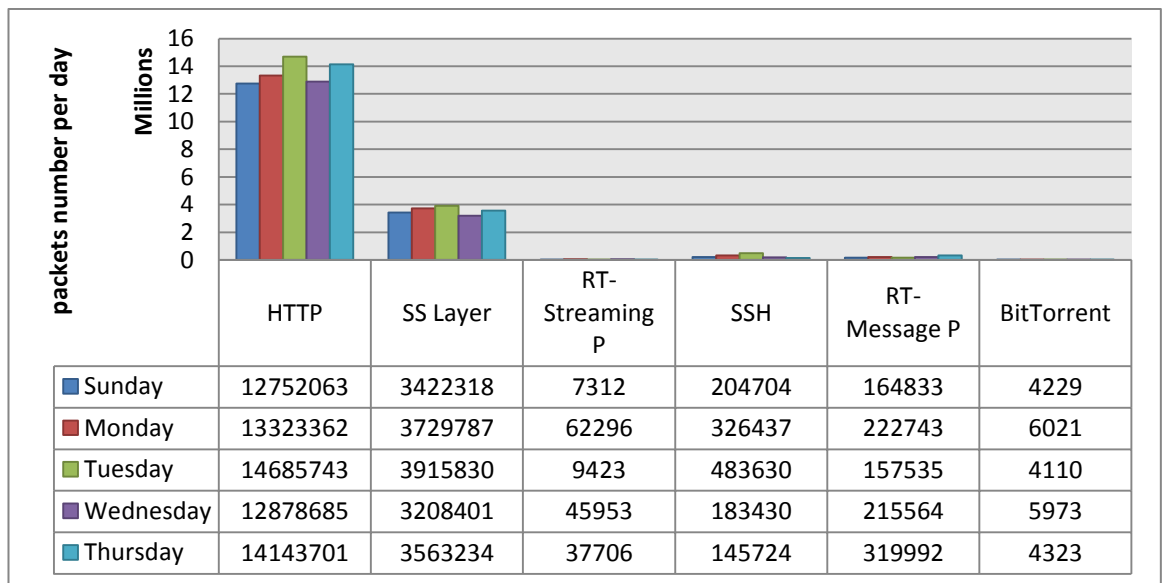
*Figure 4.8 IPv4 Distribution*

From the previous Figure, TCP protocol had a 74.5 percent higher than the rest of the protocols, because of the enormous number of applications that uses this protocol, it is possible to measure and analysis the performance of any application by isolating or filtering the relevant packets to the application and record its impact on the network.

Many applications have been discounted because it packets appeared in certain days and didn't appear in the rest of the day or because of the slight number of packages of these applications as a result will not affect the performance of the network. The next section covers the most important applications that were taken high packets rate of TCP protocol.

## TCP Protocol Distribution

There are thousands of protocols and application uses TCP protocols to transmit the data. For example file sharing applications like UTorrent, chatting applications like Yahoo messenger, E-mail applications that uses Simple Mail Transfer Protocol (SMTP) protocol and media stream applications that uses Real Time Streaming Protocol (RTSP). Following Figure illustrates most common protocols that used TCP protocol to exchange the data in UUM network.



*Figure 4.9 TCP Distribution*

HTTP is an application protocol considered as the official protocol to transfer hypertext, from the previous Table HTTP gained most TCP packets, heights packets rate was on Tuesday and Thursday, this explain the highest rate of TCP packets in Figure 4.7, this indicates that these days there were high Internet usage from the student or staff from other days, HTTP will be analyzed in chapter five studying and analyzing users preference, on the other hand SSL has the highest packets rates on

Tuesday and Monday, SSL is cryptographic protocol that be responsible for communication security in the application layer, there were variety of application used this protocol inside campus for instance web-browsing and email.

RTS protocol is responsible for controlling media streaming applications, Monday and Wednesday gained the highest packet rates, moreover, SSH protocol which is responsible for securing data communication by creating a channel over an insecure network to exchange data, highest packet rates was on Tuesday and Monday, this explain the highest rates of TCP packets on Tuesday in Figure 4.7.

RTM protocol was developed by Micormedia, this protocol was designed for audio, video and data streaming over the Internet, Monday and Thursday took the highest rates, Bit Torrent is P2P file sharing protocol used for sharing enormous quantities of data. Monday and Wednesday took the highest rate of Bit-Torrent packets. Figure 4.10 illustrates others protocols that used TCP for transmitting the data:

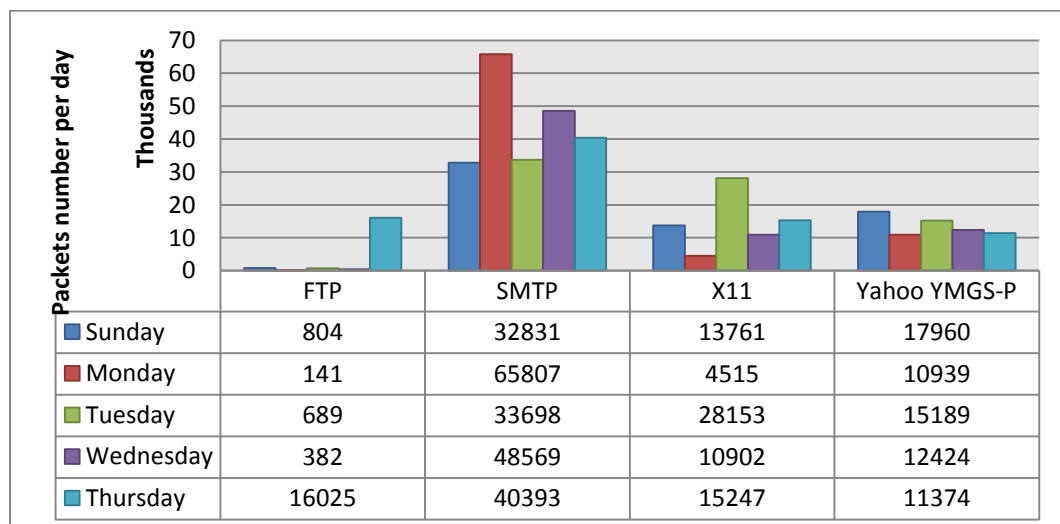


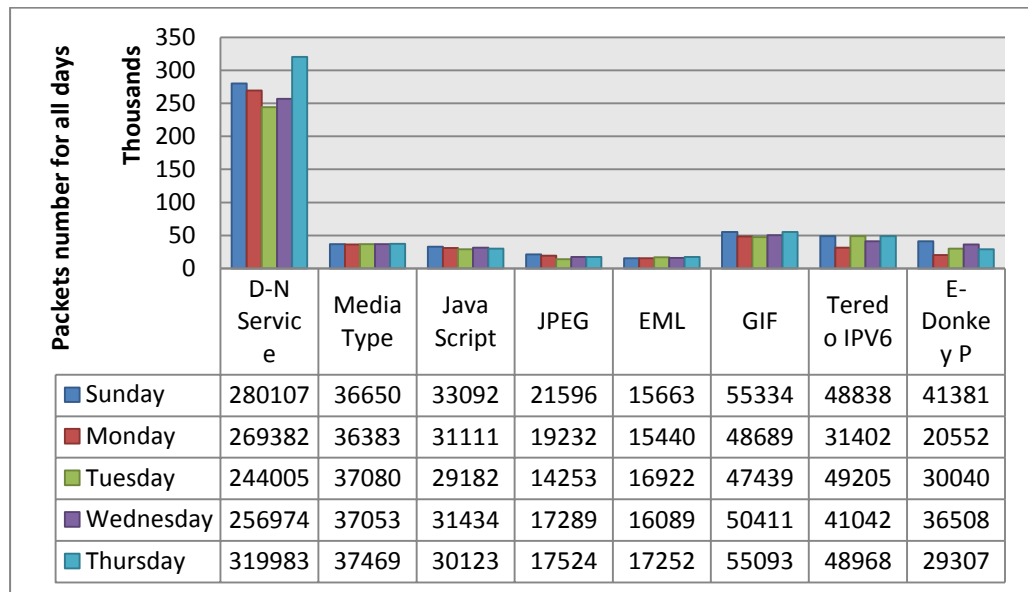
Figure 4.10 Con... TCP Distribution

FTP protocol used to transfer data between hosts over TCP protocol, from the previous Figure 4.10 FTP packet numbers were very closer except on Thursday where a high packet rate was recorded, this define the higher rates of TCP packets, Figure 4.7.

SMT protocol is an Internet standard for electronic mail E-Mail (Kurose & Ross, 2010). Highest packet rates of this protocol was recorded on Monday and Wednesday , X11 protocol offers a rich input device's ability for networked computers and GUI graphical user interfaces to communicate. Tuesday gained the highest rates. Yahoo messenger protocol supports on line and off line messaging, file transmission, voice chat and others. Sunday took the highest packet rates.

### **UDP Protocol Distribution**

This protocol occupies the second rank in order of Internet usage after the TCP protocol (Kouvatsos, 2011). “UDP is a best-effort-delivery protocol that sends data the best way it can, but doesn’t take responsibility for the data’s integrity. UDP is a store and forward protocol” (Curtis & Taylor, 2005). The following Figure describes most important protocols that use UDP to exchange data:



*Figure 4.11 UDP Distribution*

DNS protocol is exhaustive descriptions of the network data structures and communication exchanges used in domain name system, former protocol took highest rates on Thursday and Sunday, media type reflects all the photos, videos and audio that were exchanged using this protocol, which does not fall under any other protocol or media type, Thursday has the highest number of packets that used to exchange this type of media, on the other hands Java script it's scripting language used to applied as portion of any Web-browser to provide and improve user interfaces and dynamic websites. In fact Java script might be used in standalone applications such as PDF documents and desktop widgets (Flanagan, 2011). Sunday took the highest packets rates, this might be because of Sunday is the beginning day of the new working day week so all users computers start synchronizing their desktop widgets or web-browser plugins.

Joint Photographic Experts Group (JPEG) universally known as an image extension in more detail it's a method used to compress digital photography (Acharya, 2005). Sunday gained the highest packets rate of JPEG file extension.

Extensible Markup Language is a programming language that elucidates certain instructions for encoding documents in a format that both human and machine can understand (Ray, 2003), all days packets rates was closer except on Thursday where highest rate was recorded. Furthermore Graphics Interchange Format it's an image format widely used in the Internet, Sunday and Thursday had the highest rate of the GIF file format under UDP protocol.

Teredo IPv6 is a conversion technique that provides complete IPv6 specifications to hosts that using the IPv4 Internet by generating tunnels. Tuesday has the highest packets rate, E-Donkey is peer to peer file sharing protocol and Sunday took the highest packet rates of this protocol. Other UDP descried in two following Figures:

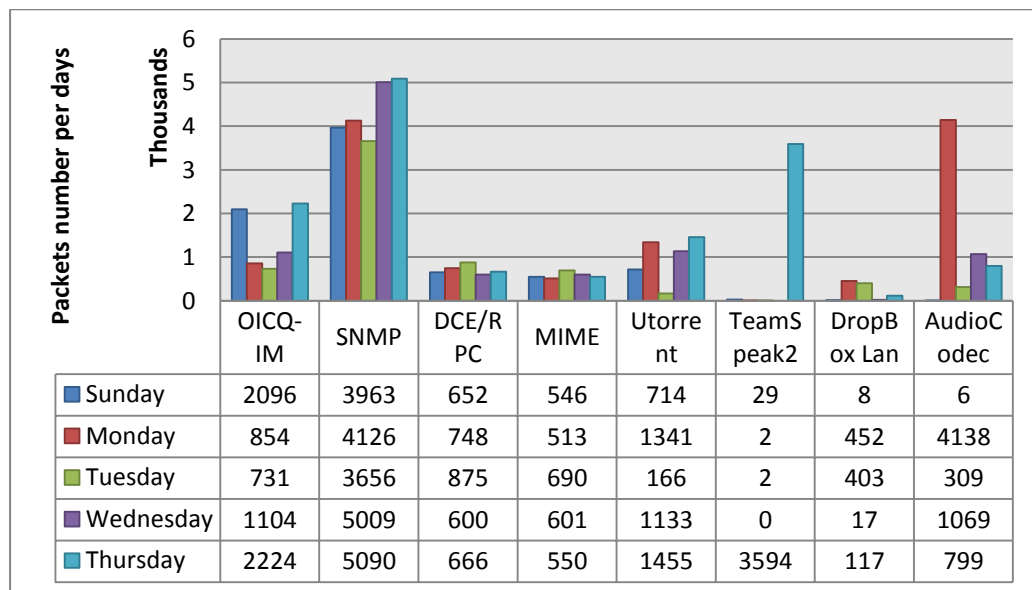


Figure 4.12 Con... UDP Distribution

OICQ is a chatting application work under ICQ instant messaging service protocol, this application is widely used in China, Sunday and Thursday took the highest number of this application packets. SNMP is a protocol for managing and controlling hosts on networks such as user's computer, printers, servers, routers. Thursday gained the highest rate.

DCE/RPC is a system established for the distributed computing environment which permits the computer programmer to code their distributed software as if it were operate on one computer. Thursday took the highest packets from other days. MIME is an Internet standard defines E-mail contents, Tuesday took the highest rate. UTorrent is a file sharing application work under UDP protocol from the previous Table Thursday also took the highest packets rate. Teamspeak2 is a chatting program that has VOIP feature, all VOIP packets was calculated with its program, a drop box is a file sharing application; on the other hand audio codec is software that implements an algorithm to compress and decompresses digital audio.

#### **4.4 Summary**

This chapter has shown the obtained results from measuring the performance of UUM network from four aspects throughput, packets loss, packets lengths and Ethernet traffic distribution, in the meantime network load was measured in addition to specifies and analyze the most used protocols by users in campus.



## **CHAPTER FIVE**

### **INTERNET USERS PREFERENCES**

#### **5.1 Introduction**

The present chapter provides elaborates on the user's web-application preferences in UUM campus by answering two questions; what web-application took most of the packets? And which countries does a highest packet go to? The answer to these questions facilitates the second objective of the study which is to determine the users' of UUM's preferred websites.

Section 5.2 provides the list of the most frequently visited sites in the UUM network, while section 5.3 provides a description and analysis of the domain-names categories to determine which of them collected the greatest number of packets, websites, and domain names were categorized according to Straubhaar & LaRose (2010) Section 5.4 provides information of which of the countries took the greatest rate of packets and finally, 5.5 contains the chapter summary.

#### **5.2 Users' Preferred Category of Websites**

Following the analysis of the entire HTTP request packets process by UUM distribution switch, and after specifying the websites traffic, it is evident that social networking sites received the highest packets request percentage compared to other sites with 42 percent of the total request packets for all the five days within an hour. It is assumed that it is owing to the social networking sites provision of video, and URL sharing devices. This is followed by search engines with 19 percent which are

indispensable to staff and student alike, and E-commerce with 9 percent. Figure 5.1 depicts the categories of websites most preferred and their percentage.

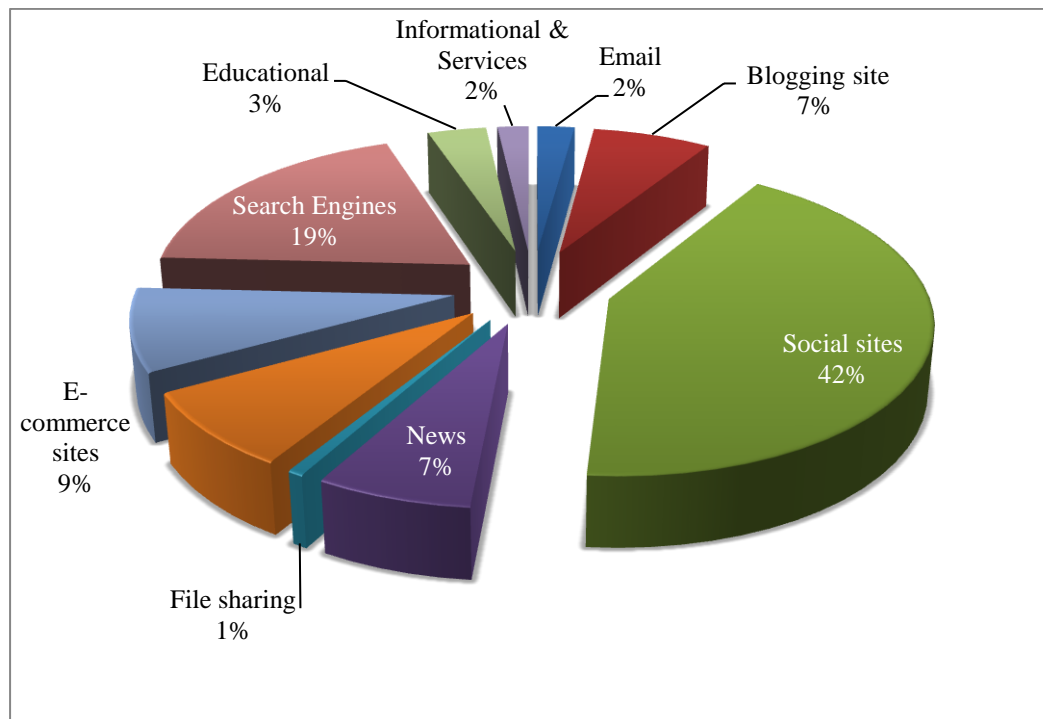


Figure 5.1: Websites Category User's Preferences

The previous Figure 5.1 demonstrates all websites categories and the percentage of each category, more information about user preferences shown in Figure 5.2 below:

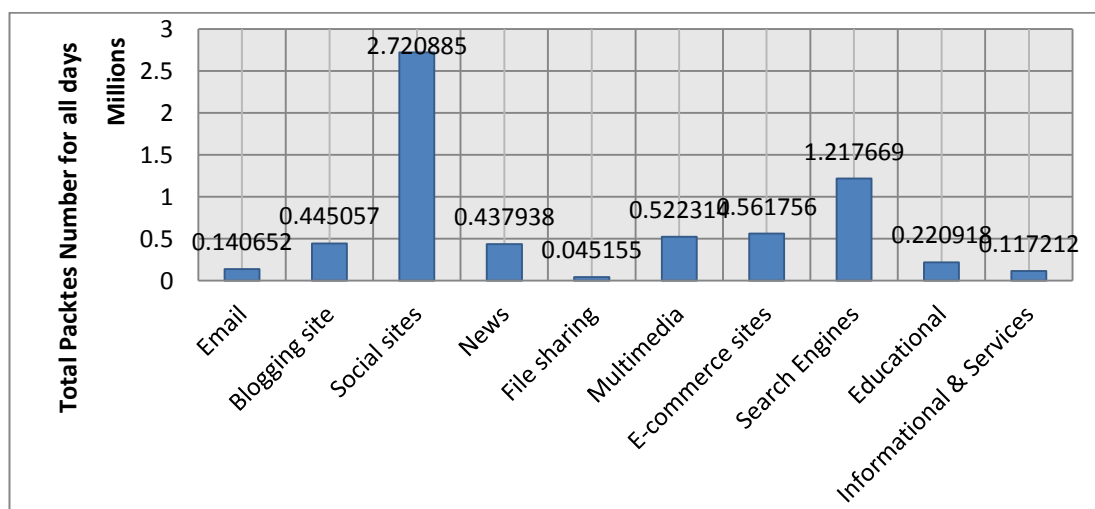


Figure 5.2: Websites Category User's Preferences Histogram

Previous Figure 5.2 categories results, the ranking is as follows: 1) Social sites 2) Search engines. 3) E-commerce websites. 4) Multimedia. 5) Blogging websites. 6) News websites. 7) Educational Websites. 8) E-mail. 9) Informational & services. 10) File sharing websites.

### **5.3 Websites Preferences**

This section will specify which sites had a high requests from users in hour each day:

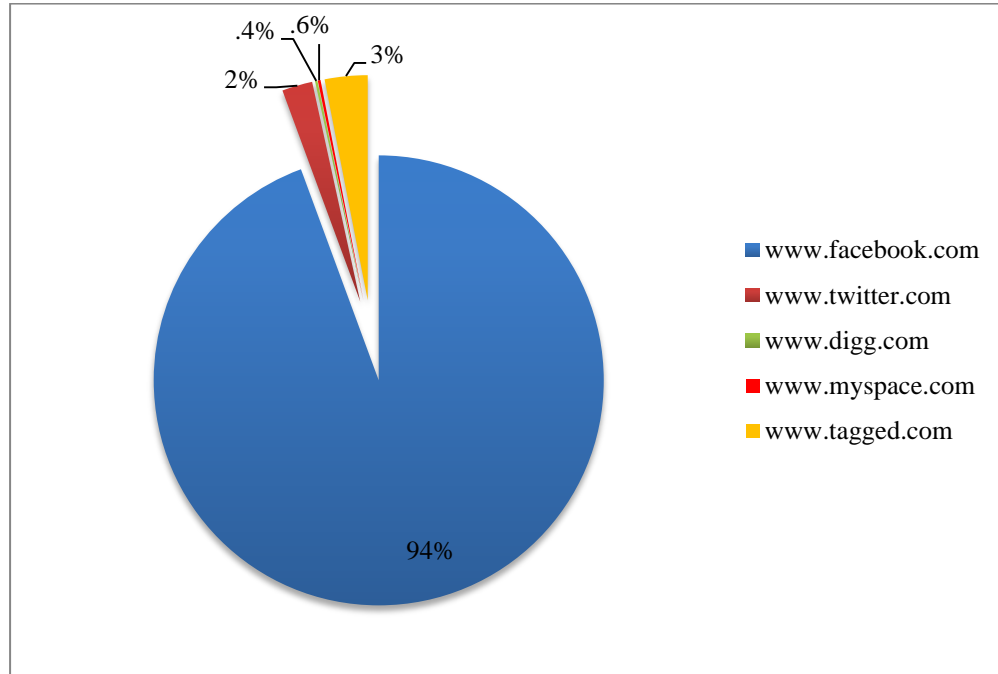
#### **5.3.1 Social Websites Preferences**

One of the advantages of social-networking sites is the ability to connect users from all over the world, whether using text, audio and video sharing these are the main reasons of the widespread usage of social-networking sites. By analyzing social site's data for all the days, it became clear that Facebook took the highest request compared to the rest of social sites. This is a normal, since Facebook is a global site, and it's one of the first sites that established the concept of sharing between users.

Some research indicates YouTube as a site for entertainment, others classify YouTube as a video sharing site and lies under this category, other researchers considered YouTube as social-networking site, the main factor to determine the classification of YouTube is the participants' number (Wang et al., 2010).

Tagged took the second rate after Facebook , after Tagged comes Twitter, Twitter took third rate, universally known as a social-networking site allow sharing information, audio and video between members. MySpace took the fourth rate and

finally comes Digg. The following Figure 5.3 presents the percentage of each social website:

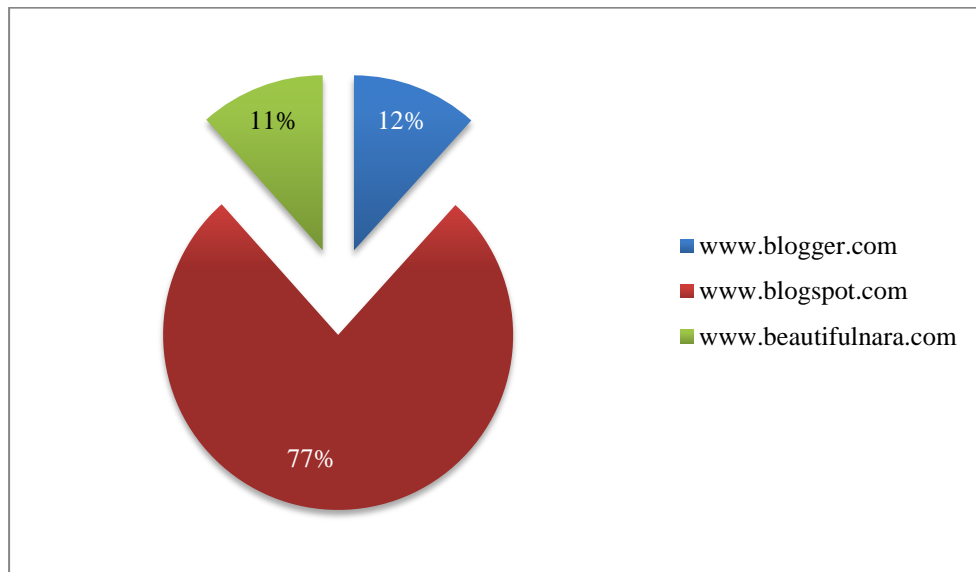


*Figure 5.3: Social Networking Sites Distribution*

Figure 5.3 demonstrate the percentage of each social website for all days, Facebook took 94 percent of the social networking sites packets, while Tagged gained 3 percent and Twitter has 2 percent, MySpace and Digg took 1 percent.

### **5.3.2 Blogs Preferences**

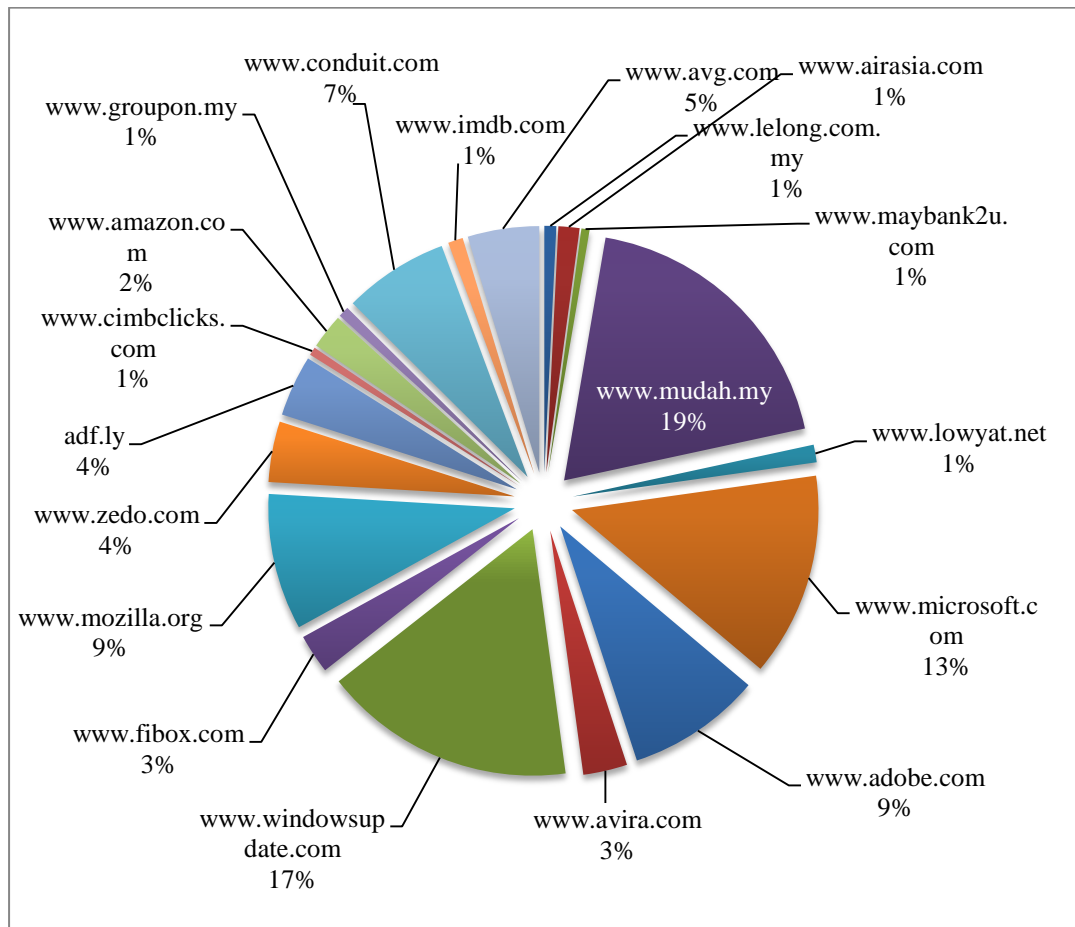
In 2007, BlogSpot and Blogger considered as most famous blogging sites worldwide because of high participant's number at these sites, after analyzing the data, BlogSpot.com took the highest rate 77 percent of packets then blogger.com with 12 percent and finally beautifulnara.com gained 11percent. Figure 5.4 presents the percentage of each website.



*Figure 5.4 Blogging websites Distribution*

### **5.3.3 E-commerce and Technical Support Websites Preferences**

Statistics for each site were mentioned on the appendix, E-commerce and technical support sites from 10:00 to 11:00 AM for each day, mudah.com website took the heights rate of requesting packets, mudah.com is one of E-commerce Malaysian websites. Next, windowsupdate.com is used for different services such as technical support and selling software, further statistic shown in Figure 5.5:



*Figure 5.5 E-commerce and Technical Support Websites Packets*

#### 5.3.4 News Websites Preferences

Figure 5.6 describes the statistic of each news website, from 10:00 to 11:00 AM daily, yahoo.com took the highest packets rate then hmetro.com has the second rate.

Further statistics shown in table and Figure 5.6 for its rate

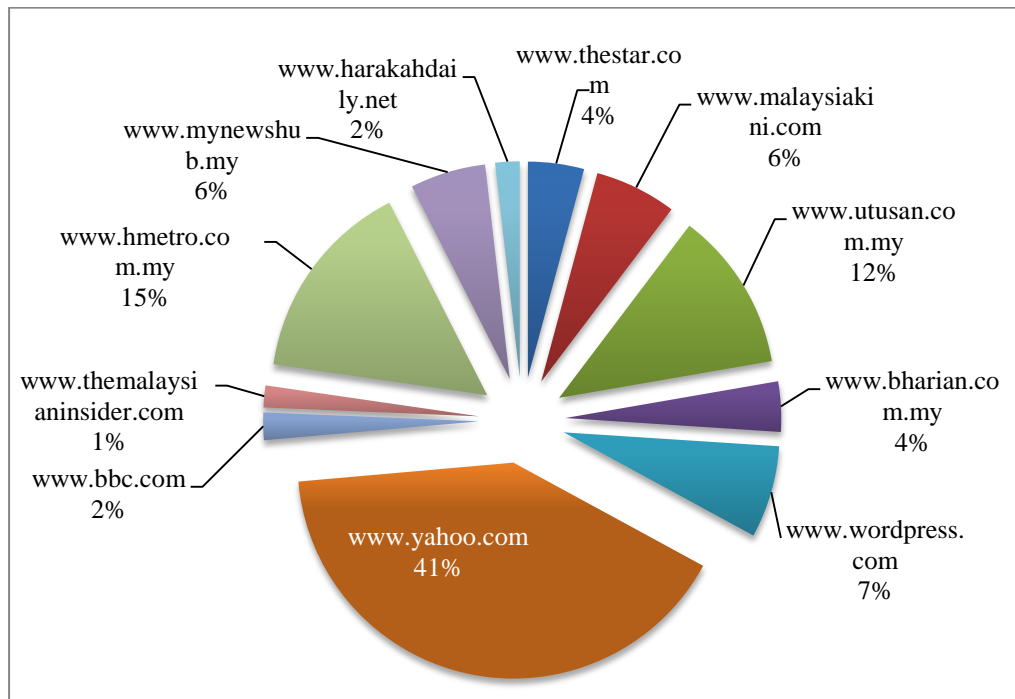


Figure 5.6 News Websites Packets

### 5.3.5 File Shearing Websites Preferences

Figure 5.7 illustrates the statistics for each file sharing website, mediafire.com and 4shared.com are well-known domain in Internet world for sharing file, 4shared.com took the highest packets rate 59 percent, and mediafire.com took 41 percent.

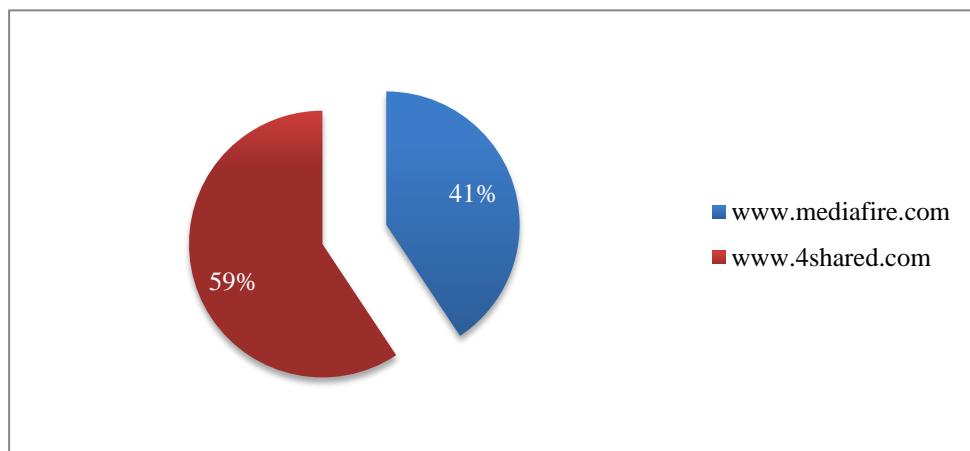
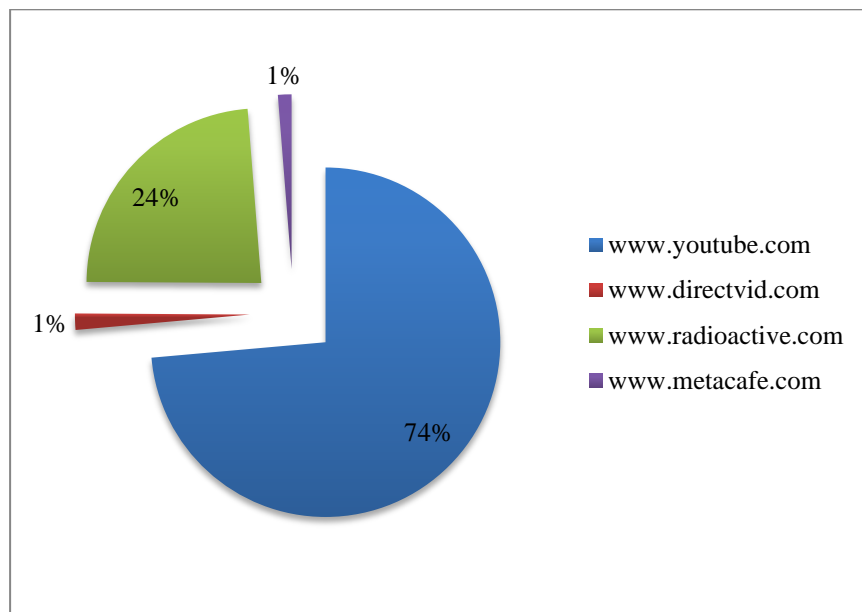


Figure 5.7 File Sharing Websites Packets

### 5.3.6 Multimedia Websites Preferences

Figure 5.8 demonstrates the statistics of each Multimedia website, YouTube took the highest rate with 74 percent then radioactive.com broadcasting radio stations through the Internet it took 24 percent of multimedia packets. Directdive.com and metacafe.com are sharing video sites took they took 2 percent.

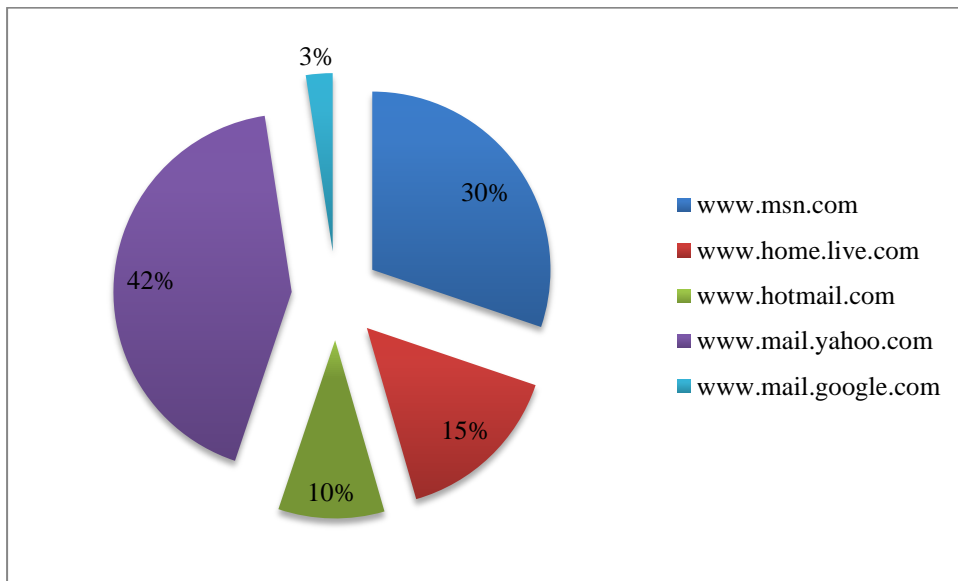


*Figure 5.8 Multimedia Websites Packets Rates*

### 5.3.7 E-mail Websites Preferences

Figure 5.9 demonstrates the statistics of each Email website from 10:00 to 11:00 AM daily, yahoo mail and messenger took the highest rate 42 percent, msn.com with msn messenger takes 30 percent, other websites statistic shown in Table and Figure 5.9 for its rate.

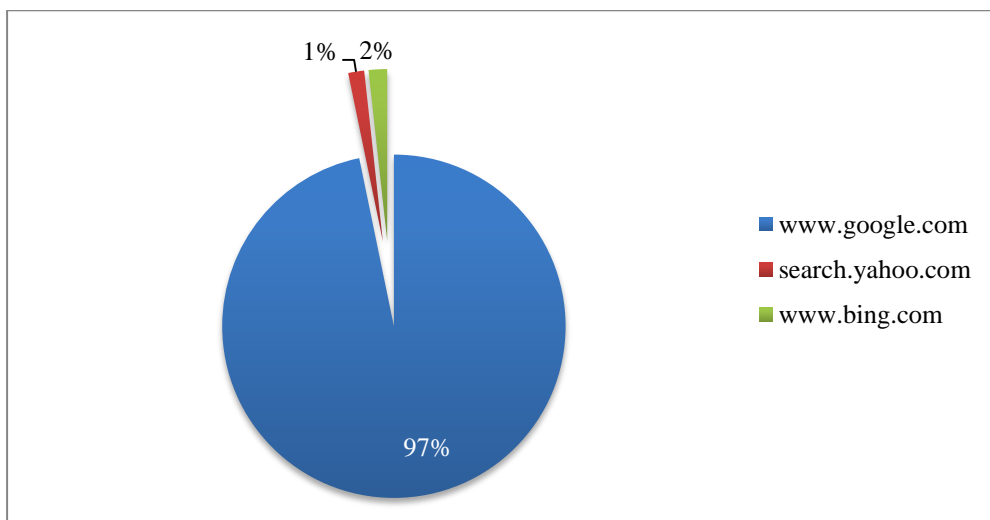




*Figure 5.9 E-mail Websites Packets Rates*

### 5.3.8 Search Engines Websites Preferences

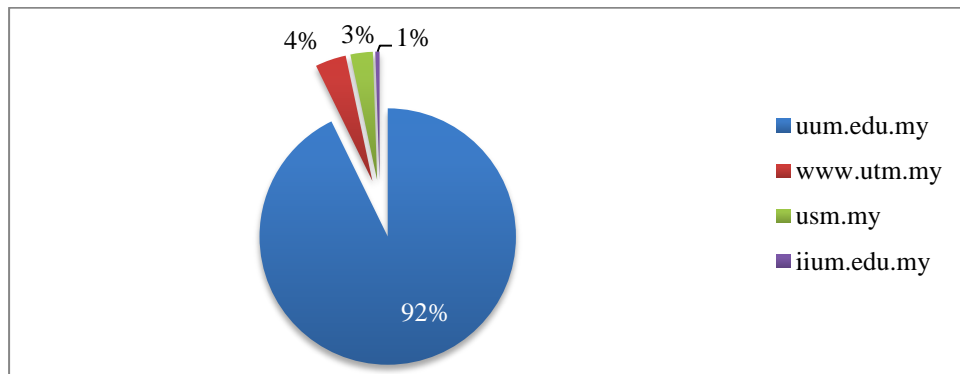
Figure 5.10 illustrates the statistics for each search engine request packets, google.com took the highest rate 97 percent, big.com took 2 percent, and search.yahoo.com took 1 percent.



*Figure 5.10 Search Engines Websites Packets Rates*

### 5.3.9 Educational Websites Preferences

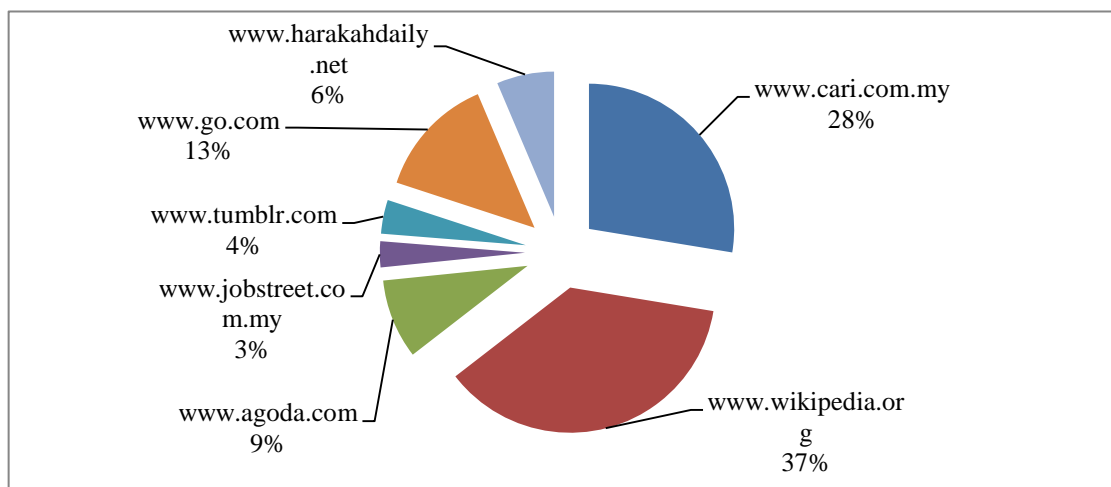
Figure 5.11 illustrates the statistic of each Educational websites request packets, uum.edu.my took the highest rate 92 percent it contain all service that fall under this domain , utm.my took 4 percent.



*Figure 5.11 Educational Engines Websites Packets Rates*

### 5.3.10 Informational and Services Websites Preferences

Figure 5.12 describes the statistics for E-commerce and technical support sites from 10:00 to 11:00 AM daily, wikipedia.org website took the heights rate of requesting packets 37 Percent, cari.com took 28 Percent.

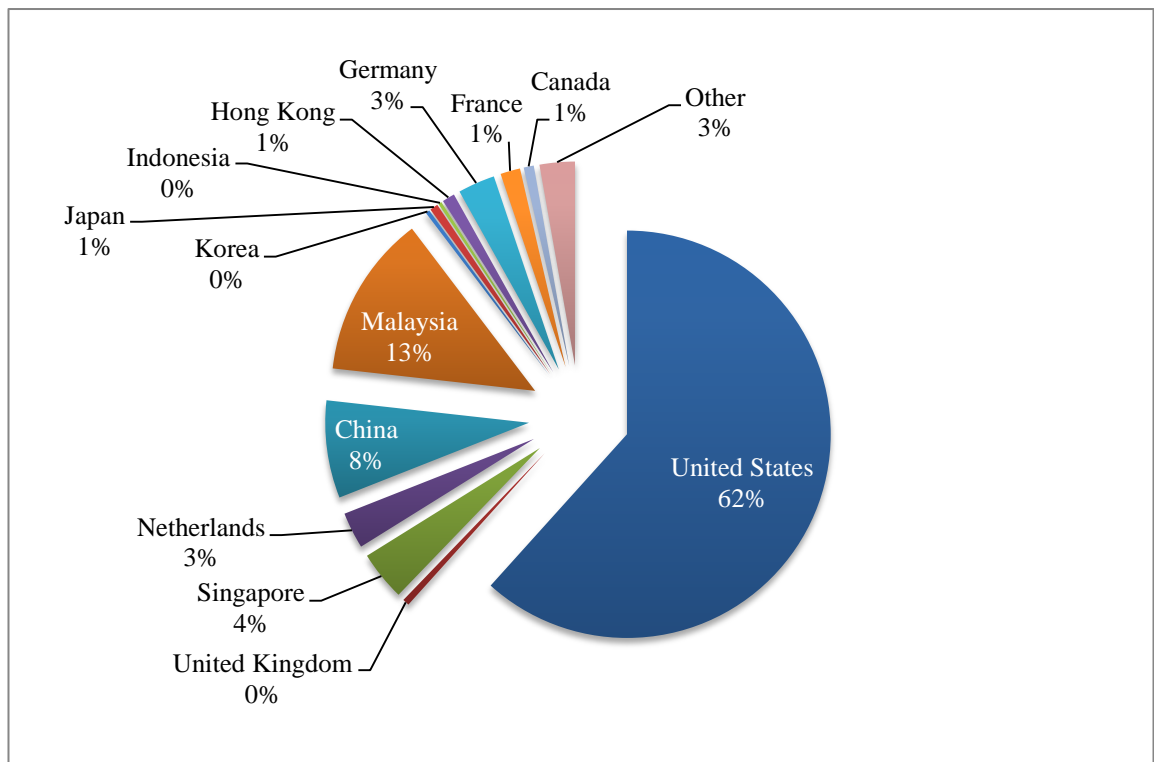


*Figure 5.12 Informational and Services packets Rates*

#### 5.4 Countries Traffic Distributions

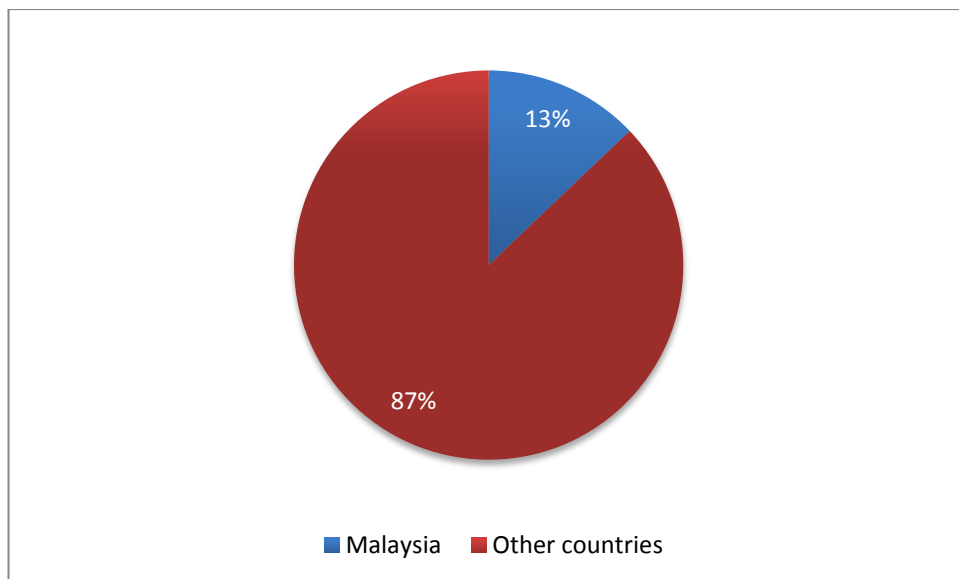
After completing the process of identifying Internet user's preferences, in UUM campus, starts new process which is to specify and determine the direction of packets, depending on GeoLite database that implemented to Wireshark software. However, this process was based on analyzing HTTP packets, whether HTTP request or HTTP responds packets.

Figure 5.13 illustrates the packets destination of several countries, United State of America has the highest packets, and this situation is typical, because main servers of most websites that has high packets rate are in USA, such as google.com, facebook.com, Twitter.com. The following Figure demonstrates HTTP packets transmission rates of each country:



*Figure 5.13 Packets Distributions over Countries*

Previous Figure illustrates the packets destination of several countries; United State of America has the highest rate with 62 percent, then Malaysia with 13 percent, third China 8 percent. Germany and Netherlands was fourth. Figure5.14 demonstrates HTTP packets transmission rates of Malaysia and international uses:



*Figure 5.14 Local and International Packets Rate*

Previous Figure illustrates the distribution of HTTP protocol, local Malaysia packets transmission was 13 percent where international transmission of the same protocol was 87 percent.

## **5.5 Summary**

The present chapter provided a review of results of the second objective of the study. First, it explains the UUM Internet users' preferences and second, it provides a description of each website that users used in an hour and finally, it discusses packet directions through many countries.

## **CHAPTER SIX**

### **FINDINGS AND FUTURE RECOMMENDATION**

#### **6.1 Introduction**

The chapter elaborates on the findings of data analysis which was discussed in the preceding chapters. It lists the problems faced during the study and its limitations. It also explains recommendations for future research.

#### **6.2 Discussion of Findings**

The present study's findings have declared certain points of interest:

With regards to the first objective, the network usage in UUM registered high rates during all the five days, with the highest usage at 23730 PPS and the lowest at 21849 PPS. These statistics' variation depends on changes in time. The results revealed that several packets were lost during transmission which required the re-transmission of the corrupted packets and resulting in the increase of transmitted packets rates in the network.

Majority of the packets were of short-length leading to exhaustion of network devices during transmission. Further studies are called for to clarify this area of finding.

The findings also revealed that the used rate of TCP and HTTP protocols was high, characterized by many errors in the TCP protocol including out-of-order segment, Zero window packets, Duplicate acknowledgement among others.

As for the second objective, through the analysis of the users' behavior, the findings revealed that the most frequently visited sites were that of social networking sites and video streaming implying the need to limit such access during working or studying hours.

### **6.3 Suggestion of Future Works**

Generally speaking, users' applications within the network must be analyzed to confine the highly-used applications like streaming applications and determine its effects of the network. Moreover, find out the reasons behind packet loss and other issues. The quality of analysis depends on the topic of study and the type of analysis tools utilized on each protocol. As well as using port filtering on distribution switches to prevent Trojans and viruses to access the network from the user's side.

Designing new system similar to internal social networking sites may allow staff and students to share information and communicate effectively as opposed to using social networking sites. Some policies such as firewall blocks in social networking sites may be modified during working or studying hours. Applying strict policies on bandwidth control manager to limit the use of non-beneficial Internet applications in the scientific field, and monitoring the network to detect new web-applications behavior.

#### **6.4 Problems and Limitation**

The time consumed in capturing packets was limited to one hour a day. Data collected from the switch had a large size of nearly 200GB and for the analysis of such data sizes; it is more effective to utilize Grid computing or servers to reduce time of analysis.

Through the use of Wireshark, Trojan, Worms, and Viruses were exposed during analysis although their statistical data was not collected as there is not official method supporting Wireshark tutorial or prior studies on how to report them.

#### **6.5 Contribution**

The study has many contributions to the field of network research. It provides an overview of the components of large networks like the UUM network comprising of switches, routers and other devices. Moreover, different types of protocols running through the network were also described. The study also provided an idea regarding the preferences that website users have in such a large network and highlighted the need for policies to control the network.

The discussed kinds of protocols will provide invaluable information to future researchers particularly those who are interested in measuring and analyzing the performance of network. The findings of the present study may also be used as reference for comparative network performance studies in the future and this may be helpful for network administrators in their plans to improve network efficiency by determining which Internet resources negatively affect the network. The findings

also provide valuable insight to network engineers in their attempts to design and improve better network performance in light of the users' behavior and requirements.

Lastly, the present study has achieved all the predetermined objectives which are: to measure UUM network performance through Internet traffic and to determine users' behavior. Recommendations for future studies and including the enhancement of network bandwidth have also been provided.



## REFERENCES

- Acharya, T. (2005). *JPEG2000 Standard for Image Compression Concepts, Algorithms and VLSI Architectures*. New Jersey, USA: Wiley Sons, INC.
- Argyrazi, K., Maniatis, P., & Singla, A. (2010). Verifiable Network-Performance Measurements. *IEEE/ACM Transactions on Networking ACM*, 19, 1224-1226.
- Augustin, B., & Mellouk, A. (2011). On Traffic Patterns of HTTP Applications. *IEEE Communications Society journal*, 11, 2-4.
- Blum, R. (2003). *Network Performance Open Source Toolkit*: Wiley Publishing, Inc., Indianapolis, Indiana.
- Canali, C., Casolari, S., & Lancellotti, R. (2010). A quantitative methodology to identify relevant users in social networks. *IEEE/ACM Transactions on Networking IEEE*, 4, 3-10.
- Cao, Y., Liu, B., & Xue, Y. (2010). Locality Analysis of BitTorrent-Like Peer-to-Peer Systems. *Communications Society journal IEEE*, 10, 1-5.
- Chang, C.-W., Huang, G., Lin, B., & Chuah, C.-N. (2011). LEISURE: A Framework for Load-Balanced Network-Wide Traffic Measurement. *IEEE Transactions on Wireless communications*, 10, 326-628.
- Chuan, X., & Hong, T. (2008). Design and complementation of a real time Traffic Measurement System in High-Speed Networks. *IEEE*.
- COMER, D. E. (2009). *Computer Networks and Internets*. New Jersey, USA: Pearson Education, Inc.
- Crandall, S., & Jasani, H. (2011). ProMix: Linux Promiscuous Wireless Packet Analysis. *IEEE Transactions on Industrial Informatics*, 1-4.

- Curtis, N., & Taylor, P. J. (2005). *Network+ ACompTIA Certification* (Fourth Edition ed.): K LLC- CompTIA.
- Donahue, G. A. (2011). *Network Warrior* (Second Edition ed.): O'Reilly.
- Dong, X., Clark, J. A., & Jacob, J. L. (2009). User Behaviour Based Phishing Websites Detection. *International Multiconference on Computer Science and Information Technology IEEE*, 6, 40-42.
- Dulaney, E., & Harwood, M. (2012). *CompTIA Network+*. USA: Paul Boger, Pearson-Inc.
- Edwards, J., & Bramante, R. (2009). *Networking Self-Teaching Guide*. Canada: Wiley Publishing, Inc.
- Fan, Z., Zhang, L., & Shen, J. (2010). A User's Preference based Method for Web Service Selection. *International Conference on Advances in Computing, Control, and Telecommunication Technologies IEEE*, 63, 784-787.
- Flanagan, D. (2011). *JavaScript: The Definitive Guide*. United States of America.: O'Reilly Media, Inc.
- Fuentetaja, I. G., & Economou, M. (2009). Analysis of users' access to museums websites. *15th International Conference on Virtual Systems and Multimedial, EEE*, 24, 123-126.
- Governor, J., Hinchcliffe, D., & Nickull, D. (2009). *Web 2.0 Architectures*. United States of America.: O'Reilly.
- Gu, T., Hong, S.-J., & Yoo, J.-B. (2009). Personal Preference for Reliable Transaction Identification on Web Service. *Seventh ACM/IEEE Symposium on Architectures for Networking and Communications Systems IEEE*, 9, 250-255.

- Gu, T., Yoo, J.-B., & Park, C.-Y. (2008). Consideration of User Preference on Internet-based Overlay Network. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 27(2), 202-203.
- Hartpence, B. (2011). *Packet Guide to Core Network Protocols* (First Edition ed.): O'Reilly Media, Inc.
- Hassan, H., Garcia, J. M., & Bockstal, C. (2009). Modeling Internet Traffic: Performance Limits. *43rd Annual IEEE/ACM International Symposium on Microarchitecture IEEE*, 3, 4-6.
- Iliofotou, M. (2009). Exploring Graph-based Network Traffic Monitoring. *Workshops of International Conference on Advanced Information Networking and Applications IEEE*, 58, 757-758.
- Jain, R., & Hassan, M. (2004). *High Performance TCP/IP Networking*: Pearson Education, Inc.
- Joseph, V., & Veciana, G. d. (2011). Stochastic Networks with Multipath Flow Control: Impact of Resource Pools on Flow-level Performance and Network Congestion. *International Conference on Control, Automation and Systems ACM* 76, 1613-1615.
- Kim, H., Claffy, k., & Fomenkov, M. (2009). Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices. *International Conference and Workshop on Emerging Trends in Technology ACM*, 32(891-893).
- Knoth, A., & Neuhäuser, D. (2010). IPv6-only Nodes in Corporate and Academic Networks. *IEEE Globecom 2010 Workshop on Heterogeneous, Multi-hop Wireless and Mobile Networks*, 11, 142-145.

- Kouvatsos, D. D. (2011). *Network Performance Engineering*. Berlin Heidelberg: Springer-Verlag.
- Kurose, J. F., & Ross, K. W. (2010). *Computer networking: a top-down approach* (5th ed ed.): Addison Wesley.
- Lammle, T. (2007). *CCNA - Cisco Certified Network Associate* (Sixth Edition ed.): Wiley Publishing, Inc., Indianapolis, Indiana.
- Lee, K., Mirchandani, D., & Zhang, X. (2010). An Investigation on Institutionalization of Websites of Firms. *15th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications IEEE*, 9, 81-87.
- Liao, Y.-W., Wang, Y.-S., & Tang, T.-I. (2011). Investigating the Influence of the Landscape Preference of Blogs, User Satisfactory and Behavioral Intention. *Eighth International Conference on Information Technology, IEEE*.
- Liu, P., & Hu, R. (2009). Research on Evaluation of E-Commerce WebSites Based on linguistic ordered weighted averaging Operator. *Workshop on Knowledge Discovery and Data Mining, IEEE*.
- Lucas, M. W. (2010). *Network Flow Analysis*. San Francisco: William Pollock, No Starch Press, Inc.
- Mahimkar, A., Song, H. H., Ge, Z., & Shaikh, A. (2010,). Detecting the Performance Impact of Upgrades in Large Operational Networks. *International Conference on Future Networks ACM*, 8, 74-67.
- Marsic, I. (2010). *computer networks, performance and quality of service*: Rutgers University.

- McFarland, S., Sambi, M., Sharma, N., & Hooda, S. (2011). *IPv6 for Enterprise Networks*. Indianapolis, IN 46240 United States of America: Cisco Press.
- Meiss, M., Menczer, F., & Vespignani, A. (March 2011). Properties and Evolution of Internet Traffic Networks from Anonymized Flow Data. *ACM Transactions on Internet Technology*, 71, 825-828.
- Mieghem, P. (2009). *Performance Analysis of Communications Networks and Systems*. New York: Cambridge University Press.
- Narayan, S., Lutui, P. R., & Vijayakumar, K. (2010). Performance Analysis of Networks with IPv4 and IPv6. *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, 6, 232-236.
- Orebaugh, A., Ramirez, G., & Burke, J. (2007). *Wireshark & Ethereal Network Protocol Analyzer Toolkit*. Syngress Publishing, Inc.
- Oskouei, R. J. (2010). Analyzing Different Aspects of Social Network Usages on Students Behaviors and Academic Performance. *3rd International Conference on Cloud Computing IEEE*, 59-60.
- Parziale, L., & Britt, D. T. (2006). *TCP/IP Tutorial and Technical Overview* (Eighth Edition ed.): IBM Corp - International Business Machines Corporation.
- Peterson, L. L., & Davie, B. S. (2012). *Computer Networks: A Systems Approach* (Fifth Edition ed.). United States of America: Elsevier, Inc.
- Qadeer, M. A., & Khan, A. H. (2010). Bottleneck Analysis and Traffic Congestion Avoidance. *ACM International Conference and Workshop on Emerging Trends in Technology*, 15, 218-219.
- Ray, E. T. (2003). *Learning XML, Second Edition*. United States of America: O'Reilly & Associates, Inc.

- Sanders, C. (2007). *practical packet analysis using wireshark to solve real-world network problems*. United States of America: William Pollock, No Starch Press, Inc.
- Shan, X., & Sun, H. (2011). The Research of Web Users' Behavior Mining Based on Association Rules. *State Natural Sciences Foundation project subsidization IEEE, 43*, 835-837.
- Sloan, J. D. (2001). *Network Troubleshooting Tools*: O'Reilly.
- Straubhaar, J., & LaRose, R. (2012). *Media Now 2012 Update* (seventh ed.). United State Of Amarica: Michael Rosenberg, Inc.
- Team, C. (2006). Router IP Traffic Export Packet Capture Enhancements. Retrieved 14th April 2012  
[http://www.cisco.com/en/US/docs/ios/12\\_4t/12\\_4t11/ht\\_rawip.pdf](http://www.cisco.com/en/US/docs/ios/12_4t/12_4t11/ht_rawip.pdf)
- Wamser, F., Pries, R., & Staehle, D. (2010). Traffic characterization of a residential wireless Internet access. *Springer Science+Business Media, LLC, 25*, 682-683.
- Wang, J. H., An, C., & Yang, J. (2010). A study of traffic, user behavior and pricing policies in a large campus network. *Symposium on Architectures for Networking and Communications Systems IEEE, 13*, 268-270.
- Wang, N. (2010). The Fuzzy Comprehensive Evaluation of User-oriented Government Websites. *IEEE Transactions on vehicular technology 59*, 1821-1823.
- Xiaojian, W. (2009). Comprehansive Evaluation on E-commerce Website Applying Improved TOPSIS Method. *International Conference on Electronic Commerce and Business Intelligence, IEEE, 11*, 119-120.

- Xu, D., Wang, S., & Yan, S. (2010). Analysis and Application of Wireshark in TCP/IP Protocol Teaching. *Fifth International Conference on Systems and Networks Communications IEEE*, 26, 381-386.
- Yang, X., Chen, X., & Jin, Y. (2011). A High-speed Real-time HTTP Performance Measurement Architecture Based on Network Processor. *IFIP International Conference on Network and Parallel Computing IEEE*, 21, 341-342.
- Yildirim, E., Suslu, I. H., & Kosar, T. (2009). Which Network Measurement Tool is Right for You? A Multidimensional Comparison Study. *9th Grid Computing Conference, IEEE*, 15, 152-155.

## APPENDIX A WEBSITES STATISTICS

### Social Websites Packets

Day Website	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.facebook.com">www.facebook.com</a>	456217	516633	668126	482196	444792	2567964
<a href="http://www.twitter.com">www.twitter.com</a>	11350	11562	8210	16825	11568	59515
<a href="http://www.digg.com">www.digg.com</a>	544	605	563	2153	738	4603
<a href="http://www.myspace.com">www.myspace.com</a>	298	875	30	3851	198	5252
<a href="http://www.tagged.com">www.tagged.com</a>	20774	9850	21637	15698	15592	83551
Packets per day	579107	628812	789213	563085	545038	

### Blogging Websites Packets

Day Website	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.blogger.com">www.blogger.com</a>	13406	7959	1965	16319	12382	52031
<a href="http://www.blogspot.com">www.blogspot.com</a>	75074	58179	93751	39182	75290	341476
<a href="http://www.beautifulnara.com">www.beautifulnara.com</a>	2154	2960	19735	20365	6336	51550
Packets per day	90634	69098	115451	75866	94008	



## E-commerce and Technical Support websites packets

Day Website	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.lelong.com.my">www.lelong.com.my</a>	1538	842	400	1120	474	4374
<a href="http://www.airasia.com">www.airasia.com</a>	1206	1907	1682	1685	1288	7768
<a href="http://www.maybank2u.com">www.maybank2u.com</a>	410	692	813	812	282	3009
<a href="http://www.mudah.my">www.mudah.my</a>	16876	13931	26185	34064	15228	106284
<a href="http://www.lowyat.net">www.lowyat.net</a>	464	1109	329	3560	1026	6488
<a href="http://www.microsoft.com">www.microsoft.com</a>	13448	15179	9632	25316	11570	75145
<a href="http://www.adobe.com">www.adobe.com</a>	8586	13196	1825	13185	12694	49486
<a href="http://www.avira.com">www.avira.com</a>	2288	5282	1430	4350	2914	16264
<a href="http://www.windowupdate.com">www.windowupdate.com</a>	16962	24539	12493	15289	23726	93009
<a href="http://www.fibox.com">www.fibox.com</a>	2172	3450	2745	2884	3216	14467
<a href="http://www.mozilla.org">www.mozilla.org</a>	13162	11985	4259	9835	11014	50255
<a href="http://www.zedo.com">www.zedo.com</a>	4064	3913	6275	4203	3900	22355
<a href="http://adf.ly">adf.ly</a>	2556	2163	1664	13753	2366	22502
<a href="http://www.cimbclicks.com">www.cimbclicks.com</a>	524	345	462	1462	240	3033
<a href="http://www.amazon.com">www.amazon.com</a>	2732	962	1637	4688	2792	12811
<a href="http://www.groupon.my">www.groupon.my</a>	596	1212	201	1421	540	3970
<a href="http://www.conduit.com">www.conduit.com</a>	6566	3564	6795	14283	7492	38700
<a href="http://www.imdb.com">www.imdb.com</a>	486	816	1823	2190	190	5505
<a href="http://www.avg.com">www.avg.com</a>	4010	700	2846	12753	6022	26331
Packets per day	98646	105787	83496	166853	106974	

## News Websites Packets

Website \ Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.thestar.com">www.thestar.com</a>	3736	3285	1637	6819	2888	18365
<a href="http://www.malaysiakini.com">www.malaysiakini.com</a>	6882	5198	3267	4153	7290	26790
<a href="http://www.utusan.com.my">www.utusan.com.my</a>	12296	9788	6485	13245	10560	52374
<a href="http://www.bharian.com.my">www.bharian.com.my</a>	4322	3871	1021	2948	4294	16456
<a href="http://www.wordpress.com">www.wordpress.com</a>	6516	2555	2659	13946	4544	30220
<a href="http://www.yahoo.com">www.yahoo.com</a>	46316	38318	26375	19635	47484	178128
<a href="http://www.bbc.com">www.bbc.com</a>	22	7511	635	835	14	9017
<a href="http://www.themalaysianinsider.com">www.themalaysianinsider.com</a>	108	1737	1637	2401	1256	7139
<a href="http://www.hmetro.com.my">www.hmetro.com.my</a>	10852	9644	13284	21791	11116	66687
<a href="http://www.mynewshub.my">www.mynewshub.my</a>	2676	180	6928	12366	2608	24758
<a href="http://www.harakahdaily.net">www.harakahdaily.net</a>	1294	840	437	4193	1240	8004
Total	95020	82927	64365	102332	93294	

## File Sharing Websites Packets

Website \ Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.mediafire.com">www.mediafire.com</a>	3736	3285	1637	6819	2888	18365
<a href="http://www.4shared.com">www.4shared.com</a>	6882	5198	3267	4153	7290	26790
Total	10618	8483	4904	10972	10178	

## Multimedia Websites Packets

Website \ Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.youtube.com">www.youtube.com</a>	89924	89287	90647	42362	72150	384370
<a href="http://www.directvid.com">www.directvid.com</a>	32	4163	826	2637	218	7876
<a href="http://www.radioactive.com">www.radioactive.com</a>	25330	27541	19674	26312	24654	123511
<a href="http://www.metacafe.com">www.metacafe.com</a>	282	1771	23	4203	278	6557
Total	115568	122762	111170	75514	97300	

## E-mail Websites Packets

Website \ Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.msn.com">www.msn.com</a>	8560	6168	12857	6451	8446	42482
<a href="http://www.home.live.com">www.home.live.com</a>	3862	5359	1764	6785	3788	21558
<a href="http://www.hotmail.com">www.hotmail.com</a>	1548	1658	6495	1825	2012	13538
<a href="http://www.mail.yahoo.com">www.mail.yahoo.com</a>	13640	15408	642	14351	15624	59665
<a href="http://www.mail.google.com">www.mail.google.com</a>	174	180	1276	1633	146	3409
Total	27784	28773	23034	31045	30016	

## Search Engines Websites Packets

Website \ Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.google.com">www.google.com</a>	250762	170082	183429	321726	252226	1178225
<a href="http://search.yahoo.com">search.yahoo.com</a>	4630	5859	1524	2021	4226	18260
<a href="http://www.bing.com">www.bing.com</a>	4772	1838	2635	7635	4304	21184
Total	260164	177779	187588	331382	260756	

### Educational Websites Packets

Day Website	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://uum.edu.my">uum.edu.my</a>	38250	43871	12853	73619	36326	204919
<a href="http://www.utm.my">www.utm.my</a>	428	1111	1688	4829	546	8602
<a href="http://usm.my">usm.my</a>	560	1462	1123	932	2140	6217
<a href="http://iiu.edu.my">iiu.edu.my</a>	70	234	219	362	295	1180
Total	39308	46678	15883	79742	39307	220918

### Informational & Services Websites Packets

Day Website	Sunday	Monday	Tuesday	Wednesday	Thursday	Websites packets
<a href="http://www.cari.com.my">www.cari.com.my</a>	7364	4153	7563	9624	5852	34556
<a href="http://www.wikipedia.org">www.wikipedia.org</a>	5324	3220	14924	18769	3996	46233
<a href="http://www.agoda.com">www.agoda.com</a>	1154	1377	1552	6018	994	11095
<a href="http://www.jobstreet.com.my">www.jobstreet.com.my</a>	778	752	937	603	552	3622
<a href="http://www.tumblr.com">www.tumblr.com</a>	728	1218	926	830	1014	4716
<a href="http://www.go.com">www.go.com</a>	1738	3226	1112	9354	1560	16990
Total	17086	13946	27014	45198	13968	117212

## Countries Traffic Distributions

Country	Sunday	Monday	Tuesday	Wednesday	Thursday	Total days
United States	8508675	7784360	7880964	8058620	8723528	40956147
United Kingdom	80052	62840	51807	84260	75262	354221
Singapore	689241	592880	466494	344240	515372	2608227
Netherlands	205800	706180	376341	262260	365024	1915605
China	1018101	969900	1114029	998100	1042558	5142688
Malaysia	1442217	1566880	1855665	1770440	1932678	8567880
Korea	41979	43540	39795	68980	20438	214732
Japan	96453	64640	111741	117580	40216	430630
Indonesia	8442	35360	41916	80500	10120	176338
Hong Kong	97104	134640	154140	101620	171380	658884
Germany	161658	579280	443982	443020	324632	1952572
France	92253	97620	348432	320200	185614	1044119
Canada	87507	97560	121338	104160	121242	531807
Other	231254	589806	318842	127179	589105	1856186
Total per day	12760736	13325486	13325486	12881159	14117169	



**UUM**  
Universiti Utara Malaysia

NetS/UM/12/012

College of Arts and Sciences  
Universiti Utara Malaysia  
06010 UUM Sintok  
Kedah Darul Aman, Malaysia  
Tel: (604) 928 6777  
Faks: (604) 928 6783  
<http://www.cas.uum.edu.my>

6 March 2012

Prof. Dr. Zulkhairi Md Dahalin  
Director of Computer Center  
Universiti Utara Malaysia

Assalamualaikum wr. wbkth.  
Dear Prof,


#### DATA COLLECTION PHASE

Regarding on the above matter, **Mr. Wisam Dawood (808266)** and **Mr. Mustafa M. H. Ibrahim (808988)** are our Msc.IT students. They are currently doing their Master project on network performance measurements. They need to capture real data of our UUM network.

In accordance with that, I apply to seek your support and help to ease their works.

Consideration and The Honourable's support, I precede with a thousand thanks.

Yours Faithfully,

  
(Dr. Mohd Hasbullah Omar)  
Head of Computer Science Department  
School of Computing  
UUM College of Arts and Sciences

TERIMA

14 MAR 2012

NetS  
Pusat Komputer  
Universiti Utara Malaysia

En. Khalil

U 7/3

En. Ali  
Tg. En. Lank  
Jln. Berkestan

113/3

TERIMA

- 7 MAR 2012

Pengarah  
Pusat Komputer  
Universiti Utara Malaysia



Universiti Pengurusan Terkemuka